

A2



PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁶ : C12N 15/12, C07K 14/47	A2	(11) International Publication Number: WO 99/06548 (43) International Publication Date: 11 February 1999 (11.02.99)
(21) International Application Number: PCT/IB98/01222 (22) International Filing Date: 31 July 1998 (31.07.98) (30) Priority Data: 08/905,135 1 August 1997 (01.08.97) US (71) Applicant (for all designated States except US): GENSET [FR/FR]; 24, rue Royale, F-75008 Paris (FR). (72) Inventors; and (75) Inventors/Applicants (for US only): DUMAS MILNE ED- WARDS, Jean-Baptiste [FR/FR]; 8, rue Grégoire-de-Tours, F-75006 Paris (FR). DUCLERT, Aymeric [FR/FR]; 6 ter, rue Victorine, F-94100 Saint-Maur (FR). LACROIX, Bruno [FR/FR]; 93, route de Vourles, F-69230 Saint-Genis Laval (FR). (74) Agents: MARTIN, Jean-Jacques et al.; Cabinet Régimbeau, 26, Avenue Kléber, F-75116 Paris (FR).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: 5'ESTs FOR NON TISSUE SPECIFIC SECRETED PROTEINS (57) Abstract <p>The sequences of 5'ESTs derived from mRNAs encoding secreted proteins are disclosed. The 5'ESTs may be to obtain cDNAs and genomic DNAs corresponding to the 5'ESTs. The 5'ESTs may also be used in diagnostic, forensic, gene therapy, and chromosome mapping procedures. Upstream regulatory sequences may also be obtained using the 5'ESTs. The 5'ESTs may also be used to design expression vectors and secretion vectors.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

5' ESTs FOR NON TISSUE SPECIFIC SECRETED PROTEINS

Background of the Invention

The estimated 50,000-100,000 genes scattered along the human chromosomes offer tremendous promise for the understanding, diagnosis, and treatment of human diseases. In addition, probes capable of specifically hybridizing to loci distributed throughout the human genome find applications in the construction of high resolution chromosome maps and in the identification of individuals.

In the past, the characterization of even a single human gene was a painstaking process, requiring years of effort. Recent developments in the areas of cloning vectors, DNA sequencing, and computer technology have merged to greatly accelerate the rate at which human genes can be isolated, sequenced, mapped, and characterized. Cloning vectors such as yeast artificial chromosomes (YACs) and bacterial artificial chromosomes (BACs) are able to accept DNA inserts ranging from 300 to 1000 kilobases (kb) or 100-400 kb in length respectively, thereby facilitating the manipulation and ordering of DNA sequences distributed over great distances on the human chromosomes. Automated DNA sequencing machines permit the rapid sequencing of human genes. Bioinformatics software enables the comparison of nucleic acid and protein sequences, thereby assisting in the characterization of human gene products.

Currently, two different approaches are being pursued for identifying and characterizing the genes distributed along the human genome. In one approach, large fragments of genomic DNA are isolated, cloned, and sequenced. Potential open reading frames in these genomic sequences are identified using bioinformatics software. However, this approach entails sequencing large stretches of human DNA which do not encode proteins in order to find the protein encoding sequences scattered throughout the genome. In addition to requiring extensive sequencing, the bioinformatics software may mischaracterize the genomic sequences obtained. Thus, the software may produce false positives in which non-coding DNA is mischaracterized as coding DNA or false negatives in which coding DNA is mislabeled as non-coding DNA.

An alternative approach takes a more direct route to identifying and characterizing human genes. In this approach, complementary DNAs (cDNAs) are synthesized from isolated messenger RNAs (mRNAs) which encode human proteins. Using this approach,

sequencing is only performed on DNA which is derived from protein coding portions of the genome. Often, only short stretches of the cDNAs are sequenced to obtain sequences called expressed sequence tags (ESTs). The ESTs may then be used to isolate or purify extended cDNAs which include sequences adjacent to the EST sequences. The extended cDNAs may contain all of the sequence of the EST which was used to obtain them or only a portion of the sequence of the EST which was used to obtain them. In addition, the extended cDNAs may contain the full coding sequence of the gene from which the EST was derived or, alternatively, the extended cDNAs may include portions of the coding sequence of the gene from which the EST was derived. It will be appreciated that there may be several extended cDNAs which include the EST sequence as a result of alternate splicing or the activity of alternative promoters.

In the past, these short EST sequences were often obtained from oligo-dT primed cDNA libraries. Accordingly, they mainly corresponded to the 3' untranslated region of the mRNA. In part, the prevalence of EST sequences derived from the 3' end of the mRNA is a result of the fact that typical techniques for obtaining cDNAs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs. (Adams *et al.*, *Nature* 377:3-174, 1996; Hillier *et al.*, *Genome Res.* 6:807-828, 1996).

In addition, in those reported instances where longer cDNA sequences have been obtained, the reported sequences typically correspond to coding sequences and do not include the full 5' untranslated region of the mRNA from which the cDNA is derived. Such incomplete sequences may not include the first exon of the mRNA, particularly in situations where the first exon is short. Furthermore, they may not include some exons, often short ones, which are located upstream of splicing sites. Thus, there is a need to obtain sequences derived from the 5' ends of mRNAs.

While many sequences derived from human chromosomes have practical applications, approaches based on the identification and characterization of those chromosomal sequences which encode a protein product are particularly relevant to diagnostic and therapeutic uses. Of the 50,000-100,000 protein coding genes, those genes encoding proteins which are secreted from the cell in which they are synthesized, as well as the secreted proteins themselves, are particularly valuable as potential therapeutic agents. Such proteins are often

involved in cell to cell communication and may be responsible for producing a clinically relevant response in their target cells.

In fact, several secretory proteins, including tissue plasminogen activator, G-CSF, GM-CSF, erythropoietin, human growth hormone, insulin, interferon- α , interferon- β ,
5 interferon- γ , and interleukin-2, are currently in clinical use. These proteins are used to treat a wide range of conditions, including acute myocardial infarction, acute ischemic stroke, anemia, diabetes, growth hormone deficiency, hepatitis, kidney carcinoma, chemotherapy induced neutropenia and multiple sclerosis. For these reasons, extended cDNAs encoding secreted proteins or portions thereof represent a particularly valuable source of therapeutic
10 agents. Thus, there is a need for the identification and characterization of secreted proteins and the nucleic acids encoding them.

In addition to being therapeutically useful themselves, secretory proteins include short peptides, called signal peptides, at their amino termini which direct their secretion. These signal peptides are encoded by the signal sequences located at the 5' ends of the coding
15 sequences of genes encoding secreted proteins. Because these signal peptides will direct the extracellular secretion of any protein to which they are operably linked, the signal sequences may be exploited to direct the efficient secretion of any protein by operably linking the signal sequences to a gene encoding the protein for which secretion is desired. In addition, portions of signal sequences may also be used to direct the intracellular import of a peptide or protein
20 of interest. This may prove beneficial in gene therapy strategies in which it is desired to deliver a particular gene product to cells other than the cell in which it is produced. Signal sequences encoding signal peptides also find application in simplifying protein purification techniques. In such applications, the extracellular secretion of the desired protein greatly facilitates purification by reducing the number of undesired proteins from which the desired
25 protein must be selected. Thus, there exists a need to identify and characterize the 5' portions of the genes for secretory proteins which encode signal peptides.

Public information on the number of human genes for which the promoters and upstream regulatory regions have been identified and characterized is quite limited. In part, this may be due to the difficulty of isolating such regulatory sequences. Upstream regulatory
30 sequences such as transcription factor binding sites are typically too short to be utilized as probes for isolating promoters from human genomic libraries. Recently, some approaches

have been developed to isolate human promoters. One of them consists of making a CpG island library (Cross, *et al.*, *Nature Genetics* 6: 236-244, 1994). The second consists of isolating human genomic DNA sequences containing SpeI binding sites by the use of SpeI binding protein. (Mortlock *et al.*, *Genome Res.* 6:327-335, 1996). Both of these approaches
5 have their limits due to a lack of specificity or of comprehensiveness.

The present 5' ESTs may be used to efficiently identify and isolate upstream regulatory regions which control the location, developmental stage, rate, and quantity of protein synthesis, as well as the stability of the mRNA. (Theil, *BioFactors* 4:87-93, 1993). Once identified and characterized, these regulatory regions may be utilized in gene therapy or
10 protein purification schemes to obtain the desired amount and locations of protein synthesis or to inhibit, reduce, or prevent the synthesis of undesirable gene products.

In addition, ESTs containing the 5' ends of secretory protein genes may include sequences useful as probes for chromosome mapping and the identification of individuals. Thus, there is a need to identify and characterize the sequences upstream of the 5' coding
15 sequences of genes encoding secretory proteins.

Summary of the Invention

The present invention relates to purified, isolated, or recombinant ESTs which include sequences derived from the authentic 5' ends of their corresponding mRNAs. The term
20 "corresponding mRNA" refers to the mRNA which was the template for the cDNA synthesis which produced the 5' EST. These sequences will be referred to hereinafter as "5' ESTs." As used herein, the term "purified" does not require absolute purity; rather, it is intended as a relative definition. Individual 5' EST clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from these
25 clones could not be obtained directly either from the library or from total human DNA. The cDNA clones are not naturally occurring as such, but rather are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The conversion of mRNA into a cDNA library involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection. Thus,
30 creating a cDNA library from messenger RNA and subsequently isolating individual clones from that library results in an approximately 10^4 - 10^6 fold purification of the native message.

Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

As used herein, the term "isolated" requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide present in a living animal is not isolated, but the same polynucleotide, separated from some or all of the coexisting materials in the natural system, is isolated.

As used herein, the term "recombinant" means that the 5' EST is adjacent to "backbone" nucleic acid to which it is not adjacent in its natural environment. Additionally, to be "enriched" the 5' ESTs will represent 5% or more of the number of nucleic acid inserts in a population of nucleic acid backbone molecules. Backbone molecules according to the present invention include nucleic acids such as expression vectors, self-replicating nucleic acids, viruses, integrating nucleic acids, and other vectors or nucleic acids used to maintain or manipulate a nucleic acid insert of interest. Preferably, the enriched 5' ESTs represent 15% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. More preferably, the enriched 5' ESTs represent 50% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules. In a highly preferred embodiment, the enriched 5' ESTs represent 90% or more of the number of nucleic acid inserts in the population of recombinant backbone molecules.

"Stringent", moderate," and "low" hybridization conditions are as defined in Example 29.

Unless otherwise indicated, a "complementary" sequence is fully complementary.

Thus, 5' ESTs in cDNA libraries in which one or more 5' ESTs make up 5% or more of the number of nucleic acid inserts in the backbone molecules are "enriched recombinant 5' ESTs" as defined herein. Likewise, 5' ESTs in a population of plasmids in which one or more 5' EST of the present invention have been inserted such that they represent 5% or more of the number of inserts in the plasmid backbone are "enriched recombinant 5' ESTs" as defined herein. However, 5' ESTs in cDNA libraries in which 5' ESTs constitute less than 5% of the number of nucleic acid inserts in the population of backbone molecules, such as libraries in

which backbone molecules having a 5' EST insert are extremely rare, are not "enriched recombinant 5' ESTs."

In particular, the present invention relates to 5' ESTs which are derived from genes encoding secreted proteins. As used herein, a "secreted" protein is one which, when
5 expressed in a suitable host cell, is transported across or through a membrane, including transport as a result of signal peptides in its amino acid sequence. "Secreted" proteins include without limitation proteins secreted wholly (e.g. soluble proteins), or partially (e.g. receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins which are transported across the membrane of the endoplasmic reticulum.

10 Such 5' ESTs include nucleic acid sequences, called signal sequences, which encode signal peptides which direct the extracellular secretion of the proteins encoded by the genes from which the 5' ESTs are derived. Generally, the signal peptides are located at the amino termini of secreted proteins.

Secreted proteins are translated by ribosomes associated with the "rough"
15 endoplasmic reticulum. Generally, secreted proteins are co-translationally transferred to the membrane of the endoplasmic reticulum. Association of the ribosome with the endoplasmic reticulum during translation of secreted proteins is mediated by the signal peptide. The signal peptide is typically cleaved following its co-translational entry into the endoplasmic reticulum. After delivery to the endoplasmic reticulum, secreted proteins may proceed through the
20 Golgi apparatus. In the Golgi apparatus, the proteins may undergo post-translational modification before entering secretory vesicles which transport them across the cell membrane.

The 5' ESTs of the present invention have several important applications. For example, they may be used to obtain and express cDNA clones which include the full protein
25 coding sequences of the corresponding gene products, including the authentic translation start sites derived from the 5' ends of the coding sequences of the mRNAs from which the 5' ESTs are derived. These cDNAs will be referred to hereinafter as "full length cDNAs." These cDNAs may also include DNA derived from mRNA sequences upstream of the translation start site. The full length cDNA sequences may be used to express the proteins
30 corresponding to the 5' ESTs. As discussed above, secreted proteins are therapeutically important. Thus, the proteins expressed from the cDNAs may be useful in treating or

controlling a variety of human conditions. The 5' ESTs may also be used to obtain the corresponding genomic DNA. The term "corresponding genomic DNA" refers to the genomic DNA which encodes the mRNA from which the 5' EST was derived.

Alternatively, the 5' ESTs may be used to obtain and express extended cDNAs
5 encoding portions of the secreted protein. The portions may comprise the signal peptides of the secreted proteins or the mature proteins generated when the signal peptide is cleaved off. The portions may also comprise polypeptides having at least 10 consecutive amino acids encoded by the extended cDNAs or full length cDNAs. Alternatively, the portions may comprise at least 15 consecutive amino acids encoded by the extended cDNAs or full length
10 cDNAs. In some embodiments, the portions may comprise at least 25 consecutive amino acids encoded by the extended cDNAs or full length cDNAs. In other embodiments, the portions may comprise at least 40 amino acids encoded by the extended cDNAs or full length cDNAs.

Antibodies which specifically recognize the entire secreted proteins encoded by the
15 extended cDNAs, full length cDNAs, or fragments thereof having at least 10 consecutive amino acids, at least 15 consecutive amino acids, at least 25 consecutive amino acids, or at least 40 consecutive amino acids may also be obtained as described below. Antibodies which specifically recognize the mature protein generated when the signal peptide is cleaved may also be obtained as described below. Similarly, antibodies which specifically recognize the
20 signal peptides encoded by the extended cDNAs or full length cDNAs may also be obtained.

In some embodiments, the extended cDNAs obtained using the 5' ESTs include the signal sequence. In other embodiments, the extended cDNAs obtained using the 5' ESTs may include the full coding sequence for the mature protein (*i.e.* the protein generated when the signal polypeptide is cleaved off). In addition, the extended cDNAs obtained using the 5'
25 ESTs may include regulatory regions upstream of the translation start site or downstream of the stop codon which control the amount, location, or developmental stage of gene expression.

As discussed above, secreted proteins are therapeutically important. Thus, the proteins expressed from the extended cDNAs or full length cDNAs obtained using the 5'
30 ESTs may be useful in treating or controlling a variety of human conditions.

The 5' ESTs (or cDNAs or genomic DNAs obtained therefrom) may be used in forensic procedures to identify individuals or in diagnostic procedures to identify individuals having genetic diseases resulting from abnormal expression of the genes corresponding to the 5' ESTs. In addition, the present invention is useful for constructing a high resolution map of the human chromosomes.

The present invention also relates to secretion vectors capable of directing the secretion of a protein of interest. Such vectors may be used in gene therapy strategies in which it is desired to produce a gene product in one cell which is to be delivered to another location in the body. Secretion vectors may also facilitate the purification of desired proteins.

The present invention also relates to expression vectors capable of directing the expression of an inserted gene in a desired spatial or temporal manner or at a desired level. Such vectors may include sequences upstream of the 5' ESTs, such as promoters or upstream regulatory sequences.

Finally, the present invention may also be used for gene therapy to control or treat genetic diseases. Signal peptides may also be fused to heterologous proteins to direct their extracellular secretion.

Bacterial clones containing Bluescript plasmids having inserts containing the 5' ESTs of the present invention (SEQ ID NOs: 38-291 are presently stored at 80°C in 4% (v/v) glycerol in the inventor's laboratories under the designations listed next to the SEQ ID NOs in II). The inserts may be recovered from the deposited materials by growing the appropriate clones on a suitable medium. The Bluescript DNA can then be isolated using plasmid isolation procedures familiar to those skilled in the art such as alkaline lysis minipreps or large scale alkaline lysis plasmid isolation procedures. If desired the plasmid DNA may be further enriched by centrifugation on a cesium chloride gradient, size exclusion chromatography, or anion exchange chromatography. The plasmid DNA obtained using these procedures may then be manipulated using standard cloning techniques familiar to those skilled in the art. Alternatively, a PCR can be done with primers designed at both ends of the EST insertion. The PCR product which corresponds to the 5' EST can then be manipulated using standard cloning techniques familiar to those skilled in the art.

One aspect of the present invention is a purified or isolated nucleic acid having the sequence of one of SEQ ID NOs: 38-291 or having a sequence complementary thereto. In one embodiment, the nucleic acid is recombinant.

5 Another aspect of the present invention is a purified or isolated nucleic acid comprising at least 10 consecutive bases of the sequence of one of SEQ ID NOs: 38-291 or one of the sequences complementary thereto.

Yet another aspect of the present invention is a purified or isolated nucleic acid comprising at least 15 consecutive bases of one of the sequences of SEQ ID NOs: 38-291 or one of the sequences complementary thereto. In one embodiment, the nucleic acid is
10 recombinant.

A further aspect of the present invention is a purified or isolated nucleic acid of at least 15 bases capable of hybridizing under stringent conditions to the sequence of one of SEQ ID NOs: 38-291 or one of the sequences complementary to the sequences of SEQ ID NOs: 38-291. In one embodiment, the nucleic acid is recombinant.

15 Another aspect of the present invention is a purified or isolated nucleic acid encoding a human gene product, said human gene product having a sequence partially encoded by one of the sequences of SEQ ID NO: 38-291.

Still another aspect of the present invention is a method of making a cDNA encoding a human secretory protein, said human secretory protein being partially encoded by one of
20 SEQ ID NOs 38-291, comprising the steps of contacting a collection of mRNA molecules from human cells with a primer comprising at least 15 consecutive nucleotides of a sequence complementary to one of SEQ ID NOs: 38-291; hybridizing said primer to an mRNA in said collection that encodes said protein; reverse transcribing said hybridized primer to make a first cDNA strand from said mRNA; making a second cDNA strand complementary to said first
25 cDNA strand; and isolating the resulting cDNA encoding said protein comprising said first cDNA strand and said second cDNA strand.

Another aspect of the invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-291 or a fragment thereof of at least 10 amino acids, said cDNA being
30 obtainable by the method described in the preceding paragraph. In one embodiment, the

cDNA comprises the full protein coding sequence of said protein which sequence is partially included in one of the sequences of SEQ ID NOs: 38-291.

Another aspect of the present invention is a method of making a cDNA encoding a human secretory protein that is partially encoded by one of SEQ ID NOs 38-291, comprising the steps of obtaining a cDNA comprising one of the sequences of SEQ ID NOs: 38-291; contacting said cDNA with a detectable probe comprising at least 15 consecutive nucleotides of said sequence of SEQ ID NO: 38-291 or a sequence complementary thereto under conditions which permit said probe to hybridize to said cDNA; identifying a cDNA which hybridizes to said detectable probe; and isolating said cDNA which hybridizes to said probe.

Another aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-291 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method described in the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-291.

Another aspect of the present invention is a method of making a cDNA comprising one of the sequence of SEQ ID NOs: 38-291, comprising the steps of contacting a collection of mRNA molecules from human cells with a first primer capable of hybridizing to the polyA tail of said mRNA; hybridizing said first primer to said polyA tail; reverse transcribing said mRNA to make a first cDNA strand; making a second cDNA strand complementary to said first cDNA strand using at least one primer comprising at least 15 nucleotides of one of the sequences of SEQ ID NOs 38-291; and isolating the resulting cDNA comprising said first cDNA strand and said second cDNA strand.

Another aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-291 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method described in the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-291.

In one embodiment of the method described in the two paragraphs above, the second cDNA strand is made by contacting said first cDNA strand with a first pair of primers, said

first pair of primers comprising a second primer comprising at least 15 consecutive nucleotides of one of the sequences of SEQ ID NOs 38-291 and a third primer having a sequence therein which is included within the sequence of said first primer; performing a first polymerase chain reaction with said first pair of nested primers to generate a first PCR product; contacting said first PCR product with a second pair of primers, said second pair of primers comprising a fourth primer, said fourth primer comprising at least 15 consecutive nucleotides of said sequence of one of SEQ ID NOs: 38-291, and a fifth primer, said fourth and fifth primers being capable of hybridizing to sequences within said first PCR product; and performing a second polymerase chain reaction, thereby generating a second PCR product.

One aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-291, or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-291.

Another aspect of the present invention is the method described four paragraphs above in which the second cDNA strand is made by contacting said first cDNA strand with a second primer comprising at least 15 consecutive nucleotides of the sequences of SEQ ID NOs: 38-291; hybridizing said second primer to said first strand cDNA; and extending said hybridized second primer to generate said second cDNA strand.

Another aspect of the present invention is an isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein partially encoded by one of SEQ ID NOs 38-291 or comprising a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method described in the preceding paragraph. In one embodiment, the cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-291.

Another aspect of the present invention is a method of making a protein comprising one of the sequences of SEQ ID NOs: 292-545, comprising the steps of obtaining a cDNA encoding the full protein sequence partially included in one of the sequences of sequence of SEQ ID NOs: 38-291; inserting said cDNA in an expression vector such that said cDNA is

operably linked to a promoter; introducing said expression vector into a host cell whereby said host cell produces the protein encoded by said cDNA; and isolating said protein.

Another aspect of the present invention is an isolated protein obtainable by the method described in the preceding paragraph.

5 Another aspect of the present invention is a method of obtaining a promoter DNA comprising the steps of obtaining DNAs located upstream of the nucleic acids of SEQ ID NOs: 38-291 or the sequences complementary thereto; screening said upstream DNAs to identify a promoter capable of directing transcription initiation; and isolating said DNA comprising said identified promoter. In one embodiment, the obtaining step comprises
10 chromosome walking from said nucleic acids of SEQ ID NOs: 38-291 or sequences complementary thereto. In another embodiment, the screening step comprises inserting said upstream sequences into a promoter reporter vector. In another embodiment, the screening step comprises identifying motifs in said upstream DNAs which are transcription factor binding sites or transcription start sites.

15 Another aspect of the present invention is an isolated promoter obtainable by the method described above.

Another aspect of the present invention is an isolated or purified protein comprising one of the sequences of SEQ ID NOs: 292-545.

Another aspect of the present invention is the inclusion of at least one of the
20 sequences of SEQ ID NOs: 38-291, or one of the sequences complementary to the sequences of SEQ ID NOs: 38-291, or a fragment thereof of at least 15 consecutive nucleotides in an array of discrete ESTs or fragments thereof of at least 15 nucleotides in length. In one embodiment, the array includes at least two of the sequences of SEQ ID NOs: 38-291, the sequences complementary to the sequences of SEQ ID NOs: 38-291, or fragments thereof of
25 at least 15 consecutive nucleotides. In another embodiment, the array includes at least five of the sequences of SEQ ID NOs: 38-291, the sequences complementary to the sequences of SEQ ID NOs: 38-291, or fragments thereof of at least 15 consecutive nucleotides.

Another aspect of the present invention is a promoter having a sequence selected from the group consisting of SEQ ID NOs: 31, 34, and 37.

Brief Description of the Drawings

Figure 1 is a summary of a procedure for obtaining cDNAs which have been selected to include the 5' ends of the mRNAs from which they derived.

Figure 2 shows the distribution of Von Heijne scores for 5' ESTs in each of the categories described herein and the probability that these 5' ESTs encode a signal peptide.

Figure 3 summarizes a general method used to clone and sequence extended cDNAs containing sequences adjacent to 5' ESTs.

Figure 4 (description of promoters structure isolated from SignalTag 5' ESTs) provides a schematic description of promoters isolated and the way they are assembled with the corresponding 5' tags.

Detailed Description of the Preferred Embodiment

Table IV is an analysis of the 43 amino acids located at the N terminus of all human SwissProt proteins to determine the frequency of false positives and false negatives using the techniques for signal peptide identification described herein.

Table V shows the distribution of 5' ESTs in each category described herein and the number of 5' ESTs in each category having a given minimum Von Heijne's score.

Table VI shows the distribution of 5' ESTs in each category described herein with respect to the tissue from which the 5' ESTs of the corresponding mRNA were obtained.

Table VII describes the transcription factor binding sites present in each of these promoters.

I. General Methods for Obtaining 5' ESTs derived from mRNAs with intact 5' ends

In order to obtain the 5' ESTs of the present invention, mRNAs with intact 5' ends must be obtained. Currently, there are two approaches for obtaining such mRNAs with intact 5' ends as described below: either chemical (1) or enzymatic (2).

1. Chemical Methods for Obtaining mRNAs having Intact 5' Ends

One of these approaches is a chemical modification method involving derivatization of the 5' ends of the mRNAs and selection of the derivatized mRNAs. The 5' ends of

eukaryotic mRNAs possess a structure referred to as a "cap" which comprises a guanosine methylated at the 7 position. The cap is joined to the first transcribed base of the mRNA by a 5', 5'-triphosphate bond. In some instances, the 5' guanosine is methylated in both the 2 and 7 positions. Rarely, the 5' guanosine is trimethylated at the 2, 7 and 7 positions. In the
5 chemical method for obtaining mRNAs having intact 5' ends, the 5' cap is specifically derivatized and coupled to a reactive group on an immobilizing substrate. This specific derivatization is based on the fact that only the ribose linked to the methylated guanosine at the 5' end of the mRNA and the ribose linked to the base at the 3' terminus of the mRNA, possess 2', 3'-cis diols.

10 Optionally, the 2', 3'-cis diol of the 3' terminal ribose may be chemically modified, substituted, converted, or eliminated, leaving only the ribose linked to the methylated guanosine at the 5' end of the mRNA with a 2', 3'-cis diol. A variety of techniques are available for eliminating the 2', 3'-cis diol on the 3' terminal ribose. For example, controlled
15 alkaline hydrolysis may be used to generate mRNA fragments in which the 3' terminal ribose is a 3'-phosphate, 2'-phosphate or (2', 3')-cyclophosphate. Thereafter, the fragment which includes the original 3' ribose may be eliminated from the mixture through chromatography on an oligodT column. Alternatively, a base which lacks the 2', 3'-cis diol may be added to the 3' end of the mRNA using an RNA ligase such as T4 RNA ligase. Example 1 below describes a method for ligation of a nucleoside diphosphate to the 3' end of messenger RNA.

20

EXAMPLE 1

Ligation of the Nucleoside Diphosphate pCp to the 3' End of mRNA.

One μg of RNA was incubated in a final reaction medium of 10 μl in the presence of 5 U of T₄ phage RNA ligase in the buffer provided by the manufacturer (Gibco -
25 BRL), 40 U of the RNase inhibitor RNasin (Promega) and, 2 μl of ³²pCp (Amersham #PB 10208). The incubation was performed at 37°C for 2 hours or overnight at 7-8°C.

Following modification or elimination of the 2', 3'-cis diol at the 3' ribose, the 2', 3'-cis diol present at the 5' end of the mRNA may be oxidized using reagents such as NaBH₄,
30 NaBH₃CN, or sodium periodate, thereby converting the 2', 3'-cis diol to a dialdehyde.

Example 2 describes the oxidation of the 2', 3'-cis diol at the 5' end of the mRNA with sodium periodate.

EXAMPLE 2

5 Oxidation of 2', 3'-cis diol at the 5' End of the mRNA with Sodium Periodate

0.1 OD unit of either a capped oligoribonucleotide of 47 nucleotides (including the cap) or an uncapped oligoribonucleotide of 46 nucleotides were treated as follows. The oligoribonucleotides were produced by *in vitro* transcription using the transcription kit "AmpliScribe T7" (Epicentre Technologies). As indicated below, the DNA template for the
10 RNA transcript contained a single cytosine. To synthesize the uncapped RNA, all four NTPs were included in the *in vitro* transcription reaction. To obtain the capped RNA, GTP was replaced by an analogue of the cap, m7G(5')ppp(5')G. This compound, recognized by the polymerase, was incorporated into the 5' end of the nascent transcript during the initiation of transcription but was not incorporated during the extension step. Consequently, the resulting
15 RNA contained a cap at its 5' end. The sequences of the oligoribonucleotides produced by the *in vitro* transcription reaction were:

+Cap:

5'm7GpppGCAUCCUACUCCCAUCCAAUCCACCCUAACUCCUCCCAUCUCCAC-
3' (SEQ ID NO:1)

20 -Cap:

5'-pppGCAUCCUACUCCCAUCCAAUCCACCCUAACUCCUCCCAUCUCCAC-3'
(SEQ ID NO:2)

The oligoribonucleotides were dissolved in 9 µl of acetate buffer (0.1 M sodium acetate, pH 5.2) and 3 µl of freshly prepared 0.1 M sodium periodate solution. The mixture
25 was incubated for 1 hour in the dark at 4°C or room temperature. Thereafter, the reaction was stopped by adding 4 µl of 10% ethylene glycol. The product was ethanol precipitated, resuspended in at least 10 µl of water or appropriate buffer and dialyzed against water.

The resulting aldehyde groups may then be coupled to molecules having a reactive
30 amine group, such as hydrazine, carbazide, thiocarbazide or semicarbazide groups, in order to facilitate enrichment of the 5' ends of the mRNAs. Molecules having reactive amine groups

which are suitable for use in selecting mRNAs having intact 5' ends include avidin, proteins, antibodies, vitamins, ligands capable of specifically binding to receptor molecules, or oligonucleotides. Example 3 below describes the coupling of the resulting dialdehyde to biotin.

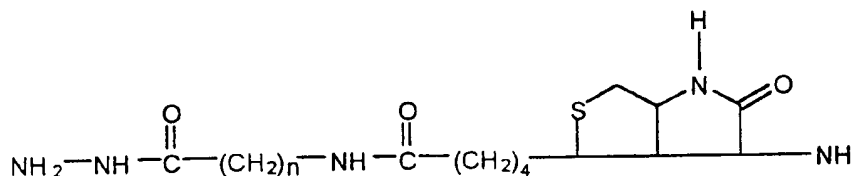
5

EXAMPLE 3

Coupling of the Dialdehyde at the 5' End of Transcripts with Biotin

The oxidation product obtained in Example 2 was dissolved in 50 μ l of sodium acetate at a pH between 5 and 5.2 and 50 μ l of freshly prepared 0.02 M solution of biotin hydrazide in a methoxyethanol/water mixture (1:1) of formula:

10



In the compound used in these experiments, $n=5$. However, it will be appreciated that other commercially available hydrazides may also be used, such as molecules of the above formula in which n varies from 0 to 5. The mixture was then incubated for 2 hours at 37°C, precipitated with ethanol and dialyzed against distilled water. Example 4 demonstrates the specificity of the biotinylation reaction.

15

20

EXAMPLE 4

Specificity of Biotinylation of Capped Transcripts

The specificity of the biotinylation for capped mRNAs was evaluated by gel electrophoresis of the following samples:

Sample 1. The 46 nucleotide uncapped *in vitro* transcript prepared as in Example 2 and labeled with 32 pCp as described in Example 1.

25

Sample 2. The 46 nucleotide uncapped *in vitro* transcript prepared as in Example 2, labeled with ^{32}pCp as described in Example 1, treated with the oxidation reaction of Example 2, and subjected to the biotinylation conditions of Example 3.

Sample 3. The 47 nucleotide capped *in vitro* transcript prepared as in Example 2 and
5 labeled with ^{32}pCp as described in Example 1.

Sample 4. The 47 nucleotide capped *in vitro* transcript prepared as in Example 2, labeled with ^{32}pCp as described in Example 1, treated with the oxidation reaction of Example 2, and subjected to the biotinylation conditions of Example 3.

Samples 1 and 2 had identical migration rates, demonstrating that the uncapped
10 RNAs were not oxidized and biotinylated. Sample 3 migrated more slowly than Samples 1 and 2, while Sample 4 exhibited the slowest migration. The difference in migration of the RNAs in Samples 3 and 4 demonstrates that the capped RNAs were specifically biotinylated.

In some cases, mRNAs having intact 5' ends may be enriched by binding the molecule
15 containing a reactive amine group to a suitable solid phase substrate such as the inside of the vessel containing the mRNAs, magnetic beads, chromatography matrices, or nylon or nitrocellulose membranes. For example, where the molecule having a reactive amine group is biotin, the solid phase substrate may be coupled to avidin or streptavidin. Alternatively, where the molecule having the reactive amine group is an antibody or receptor ligand, the
20 solid phase substrate may be coupled to the cognate antigen or receptor. Finally, where the molecule having a reactive amine group comprises an oligonucleotide, the solid phase substrate may comprise a complementary oligonucleotide.

The mRNAs having intact 5' ends may be released from the solid phase following the enrichment procedure. For example, where the dialdehyde is coupled to biotin hydrazide and
25 the solid phase comprises streptavidin, the mRNAs may be released from the solid phase by simply heating to 95 degrees Celsius in 2% SDS. In some methods, the molecule having a reactive amine group may also be cleaved from the mRNAs having intact 5' ends following enrichment. Example 5 describes the capture of biotinylated mRNAs with streptavidin coated beads and the release of the biotinylated mRNAs from the beads following enrichment.

EXAMPLE 5

Capture and Release of Biotinylated mRNAs Using Streptavidin Coated Beads

The streptavidin coated magnetic beads were prepared according to the manufacturer's instructions (CPG Inc., USA). The biotinylated mRNAs were added to a
5 hybridization buffer (1.5 M NaCl, pH 5 - 6). After incubating for 30 minutes, the unbound and nonbiotinylated material was removed. The beads were then washed several times in water with 1% SDS. The beads thus obtained were incubated for 15 minutes at 95°C in water containing 2% SDS.

Example 6 demonstrates the efficiency with which biotinylated mRNAs were
10 recovered from the streptavidin coated beads.

EXAMPLE 6

Efficiency of Recovery of Biotinylated mRNAs

The efficiency of the recovery procedure was evaluated as follows. Capped RNAs
15 were labeled with 32 pCp, oxidized, biotinylated and bound to streptavidin coated beads as described above. Subsequently, the bound RNAs were incubated for 5, 15 or 30 minutes at 95°C in the presence of 2% SDS.

The products of the reaction were analyzed by electrophoresis on 12% polyacrylamide gels under denaturing conditions (7 M urea). The gels were subjected to
20 autoradiography. During this manipulation, the hydrazone bonds were not reduced.

Increasing amounts of nucleic acids were recovered as incubation times in 2% SDS increased, demonstrating that biotinylated mRNAs were efficiently recovered.

In an alternative method for obtaining mRNAs having intact 5' ends, an
25 oligonucleotide which has been derivatized to contain a reactive amine group is specifically coupled to mRNAs having an intact cap. Preferably, the 3' end of the mRNA is blocked prior to the step in which the aldehyde groups are joined to the derivatized oligonucleotide, as described above, so as to prevent the derivatized oligonucleotide from being joined to the 3' end of the mRNA. For example, pCp may be attached to the 3' end of the mRNA using T4
30 RNA ligase as described in example 1. However, as discussed above, blocking the 3' end of

the mRNA is an optional step. Derivatized oligonucleotides may be prepared as described in Example 7.

EXAMPLE 7

5

Derivatization of Oligonucleotides

An oligonucleotide phosphorylated at its 3' end was converted to a 3' hydrazide in 3' by treatment with an aqueous solution of hydrazine or of dihydrazide of the formula $H_2N(R_1)NH_2$ at about 1 to 3 M, and at pH 4.5 at a temperature of 8°C overnight. This incubation was performed in the presence of a carbodiimide type agent soluble in water such as 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide at a final concentration of 0.3 M.

The derivatized oligonucleotide was then separated from the other agents and products using a standard technique for isolating oligonucleotides.

As discussed above, the mRNAs to be enriched may be treated to eliminate the 3' OH groups which may be present thereon. This may be accomplished by enzymatic ligation of sequences lacking a 3' OH, such as pCp, as described in Example 1. Alternatively, the 3' OH groups may be eliminated by alkaline hydrolysis as described in Example 8 below.

EXAMPLE 8

Elimination of 3' OH Groups of mRNA Using Alkaline Hydrolysis

In a total volume of 100 μ l of 0.1 N sodium hydroxide, 1.5 μ g mRNA is incubated for 40 to 60 minutes at 4°C. The solution is neutralized with acetic acid and precipitated with ethanol.

Following the optional elimination of the 3' OH groups, the diol groups at the 5' ends of the mRNAs are oxidized as described below in Example 9.

25

EXAMPLE 9

Oxidation of Diols of mRNA

Up to 1 OD unit of RNA was dissolved in 9 μ l of buffer (0.1 M sodium acetate, pH 6-7) or water and 3 μ l of freshly prepared 0.1 M sodium periodate solution. The reaction was incubated for 1 h in the dark at 4°C or room temperature. Following the incubation, the reaction was stopped by adding 4 μ l of 10% ethylene glycol. Thereafter the mixture was

30

incubated at room temperature for 15 minutes. After ethanol precipitation, the product was resuspended in at least 10 μ l of water or appropriate buffer and dialyzed against water.

Following oxidation of the diol groups at the 5' ends of the mRNAs, the derivatized oligonucleotide was joined to the resulting aldehydes as described in Example 10.

5

EXAMPLE 10

Ligature of Aldehydes of mRNA to Derivatized Oligonucleotides

The oxidized mRNA was dissolved in an acidic medium such as 50 μ l of sodium acetate pH 4-6. Fifty μ l of a solution of the derivatized oligonucleotide were added in order to obtain an mRNA:derivatized oligonucleotide ratio of 1:20. The mixture was reduced with a borohydride and incubated for 2 h at 37°C or overnight (14 h) at 10°C. The mixture was then ethanol precipitated, resuspended in 10 μ l or more of water or appropriate buffer and dialyzed against distilled water. If desired, the resulting product may be analyzed using acrylamide gel electrophoresis, HPLC analysis, or other conventional techniques.

15

Following the attachment of the derivatized oligonucleotide to the mRNAs, a reverse transcription reaction may be performed as described in Example 11 below.

EXAMPLE 11

Reverse Transcription of mRNAs Ligatured to Derivatized Oligonucleotides

An oligodeoxyribonucleotide was derivatized as follows. Three OD units of an oligodeoxyribonucleotide of sequence 5'ATCAAGAATTCGCACGAGACCATTAA3' (SEQ ID NO:3) having 5'-OH and 3'-P ends were dissolved in 70 μ l of a 1.5 M hydroxybenzotriazole solution, pH 5.3, prepared in dimethylformamide/water (75:25) containing 2 μ g of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide. The mixture was incubated for 2 h 30 min at 22°C and then precipitated twice in LiClO₄/acetone. The pellet was resuspended in 200 μ l of 0.25 M hydrazine and incubated at 8°C from 3 to 14 h. Following the hydrazine reaction, the mixture was precipitated twice in LiClO₄/acetone.

The messenger RNAs to be reverse transcribed were extracted from blocks of placenta having sides of 2 cm which had been stored at -80°C. The total RNA was extracted

30

using conventional acidic phenol techniques. Oligo-dT chromatography was used to purify the mRNAs. The integrity of the mRNAs was checked by Northern-blotting.

The diol groups on 7 μ g of the placental mRNAs were oxidized as described above in Example 9. The derivatized oligonucleotide was joined to the mRNAs as described in Example 10 above except that the precipitation step was replaced by an exclusion chromatography step to remove derivatized oligodeoxyribonucleotides which were not joined to mRNAs. Exclusion chromatography was performed as follows:

Ten ml of Ultrogel AcA34 (BioSeptra#230151) gel, a mix of agarose and acrylamide, were equilibrated in 50 ml of a solution of 10 mM Tris pH 8.0, 300 mM NaCl, 1 mM EDTA, and 0.05% SDS. The mixture was allowed to sediment. The supernatant was eliminated and the gel was resuspended in 50 ml of buffer. This procedure was repeated 2 or 3 times.

A glass bead (diameter 3 mm) was introduced into a 2 ml disposable pipette (length 25 cm). The pipette was filled with the gel suspension until the height of the gel stabilized at 1 cm from the top of the pipette. The column was then equilibrated with 20 ml of equilibration buffer (10 mM Tris HCl pH 7.4, 20 mM NaCl).

Ten μ l of the mRNA which had reacted with the derivatized oligonucleotide were mixed in 39 μ l of 10 mM urea and 2 μ l of blue-glycerol buffer, which had been prepared by dissolving 5 mg of bromophenol blue in 60% glycerol (v/v), and passing the mixture through a 0.45 μ m diameter filter.

The column was then loaded with the mRNAs coupled to the oligonucleotide. As soon as the sample had penetrated, equilibration buffer was added. Hundred μ l fractions were then collected. Derivatized oligonucleotide which had not been attached to mRNA appeared in fraction 16 and later fractions. Thus, fractions 3 to 15 were combined and precipitated with ethanol.

To determine whether the derivatized oligonucleotide was actually linked to mRNA, one tenth of the combined fractions were spotted twice on a nylon membrane and hybridized to a radioactive probe using conventional techniques. The 32 P labeled probe used in these hybridizations was an oligodeoxyribonucleotide of sequence 5'TAATGGTCTCGTGCGAATTCTTGAT3' (SEQ ID NO:4) anticomplementary to the derivatized oligonucleotide. A signal observed after autoradiography, indicated that the derivatized oligonucleotide had been truly joined to the mRNA.

The remaining nine tenth of the mRNAs which had reacted with the derivatized oligonucleotide was reverse transcribed as follows. A reverse transcription reaction was carried out with reverse transcriptase following the manufacturer's instructions and 50 pmol of nonamers with random sequence as primers.

5 To ensure that reverse transcription had been carried out through the cap structure, two types of experiments were performed.

In the first approach, after elimination of RNA of the cDNA:RNA heteroduplexes obtained from the reverse transcription reaction by an alkaline hydrolysis, a portion of the resulting single stranded cDNAs was spotted on a positively charged membrane and
10 hybridized, using conventional methods, to a ³²P labeled probe having a sequence identical to that of the derivatized oligonucleotide. Control spots containing, 1 pmol, 100 fmol, 50 fmol, 10 fmol and 1 fmol of a control oligodeoxyribonucleotide of sequence identical to that of the derivatized oligonucleotide were included. The signal observed in the spots containing the cDNA indicated that approximately 15 fmol of the derivatized oligonucleotide had been
15 reverse transcribed. These results demonstrate that the reverse transcription can be performed through the cap and, in particular, that reverse transcriptase crosses the 5'-P-P-P-5' bond of the cap of eukaryotic messenger RNAs.

In the second type of experiment, the single stranded cDNAs obtained from the above first strand synthesis were used as template for PCR reactions. Two types of reactions
20 were carried out. First, specific amplification of the mRNAs for alpha globin, dehydrogenase, pp15 and elongation factor E4 were carried out using the following pairs of oligodeoxyribonucleotide primers.

alpha-globin

25 GLO-S: 5'CCG ACA AGA CCA ACG TCA AGG CCG C3' (SEQ ID NO:5)
GLO-As: 5'TCA CCA GCA GGC AGT GGC TTA GGA G 3' (SEQ ID NO:6)

dehydrogenase

3 DH-S: 5'AGT GAT TCC TGC TAC TTT GGA TGG C3' (SEQ ID NO:7)
30 3 DH-As: 5'GCT TGG TCT TGT TCT GGA GTT TAG A3' (SEQ ID NO:8)

pp15

PP15-S: 5'TCC AGA ATG GGA GAC AAG CCA ATT T3' (SEQ ID NO:9)

PP15-As: 5'AGG GAG GAG GAA ACA GCG TGA GTC C3' (SEQ ID NO:10)

5

Elongation factor E4

EFA1-S: 5'ATG GGA AAG GAA AAG ACT CAT ATC A3' (SEQ ID NO:11)

EF1A-As: 5'AGC AGC AAC AAT CAG GAC AGC ACA G3' (SEQ ID NO:12)

10 Second, non specific amplifications were also carried out with the antisense oligodeoxyribonucleotides of the pairs described above and with a primer derived from the sequence of the derivatized oligodeoxyribonucleotide (5'ATCAAGAATTTCGCACGAGACCATT3') (SEQ ID NO:13).

One twentieth of the following RT-PCR product samples were run on a 1.5% agarose gel and stained with ethidium bromide.

15

Sample 1: The products of a PCR reaction using the globin primers of SEQ ID NOs 5 and 6 in the presence of cDNA.

Sample 2: The products of a PCR reaction using the globin primers of SEQ ID NOs 5 and 6 in the absence of added cDNA.

20 Sample 3: The products of a PCR reaction using the dehydrogenase primers of SEQ ID NOs 7 and 8 in the presence of cDNA.

Sample 4: The products of a PCR reaction using the dehydrogenase primers of SEQ ID NOs 7 and 8 in the absence of added cDNA.

Sample 5: The products of a PCR reaction using the pp15 primers of SEQ ID NOs 9 and 10 in the presence of cDNA.

25

Sample 6: The products of a PCR reaction using the pp15 primers of SEQ ID NOs 9 and 10 in the absence of added cDNA.

Sample 7: The products of a PCR reaction using the EIF4 primers of SEQ ID NOs 11 and 12 in the presence of added cDNA.

30 Sample 8: The products of a PCR reaction using the EIF4 primers of SEQ ID NOs 11 and 12 in the absence of added cDNA.

A band of the size expected for the PCR product was observed only in samples 1, 3, 5 and 7, thus indicating the presence of the corresponding sequence in the cDNA population.

5 PCR reactions were also carried out with the antisense oligonucleotides of the globin and dehydrogenase primers (SEQ ID NOs 6 and 8) and an oligonucleotide whose sequence corresponds to that of the derivatized oligonucleotide. The presence of PCR products of the expected size in the samples equivalent to above samples 1 and 3 indicated that the derivatized oligonucleotide had been linked to mRNA.

The above examples summarize the chemical procedure for enriching mRNAs for
10 those having intact 5' ends as illustrated in Figure 1. Further detail regarding the chemical approaches for obtaining such mRNAs are disclosed in International Application No. WO96/34981, published November 7, 1996, which is incorporated herein by reference. Strategies based on the above chemical modifications to the 5' cap structure may be utilized to generate cDNAs selected to include the 5' ends of the mRNAs from which they derived.
15 In one version of such procedures, the 5' ends of the mRNAs are modified as described above. Thereafter, a reverse transcription reaction is conducted to extend a primer complementary to the 5' end of the mRNA. Single stranded RNAs are eliminated to obtain a population of cDNA/mRNA heteroduplexes in which the mRNA includes an intact 5' end. The resulting heteroduplexes may be captured on a solid phase coated with a molecule
20 capable of interacting with the molecule used to derivatize the 5' end of the mRNA. Thereafter, the strands of the heteroduplexes are separated to recover single stranded first cDNA strands which include the 5' end of the mRNA. Second strand cDNA synthesis may then proceed using conventional techniques. For example, the procedures disclosed in WO 96/34981 or in Carninci. *et al.*, *Genomics* 37:327-336, 1996, the disclosures of which are
25 incorporated herein by reference, may be employed to select cDNAs which include the sequence derived from the 5' end of the coding sequence of the mRNA.

Following ligation of the oligonucleotide tag to the 5' cap of the mRNA, a reverse transcription reaction is conducted to extend a primer complementary to the mRNA to the 5' end of the mRNA. Following elimination of the RNA component of the
30 resulting heteroduplex using standard techniques, second strand cDNA synthesis is conducted with a primer complementary to the oligonucleotide tag.

2. Enzymatic Methods for Obtaining mRNAs having Intact 5' Ends

Other techniques for selecting cDNAs extending to the 5' end of the mRNA from which they are derived are fully enzymatic. Some versions of these techniques are disclosed in Dumas Milne Edwards J.B. (Doctoral Thesis of Paris VI University, Le clonage des ADNc
5 complets: difficultes et perspectives nouvelles. Apports pour l'etude de la regulation de l'expression de la tryptophane hydroxylase de rat, 20 Dec. 1993), EP0 625572 and Kato *et al.*, *Gene* 150:243-250, 1994, the disclosures of which are incorporated herein by reference.

Briefly, in such approaches, isolated mRNA is treated with alkaline phosphatase to remove the phosphate groups present on the 5' ends of uncapped incomplete mRNAs.
10 Following this procedure, the cap present on full length mRNAs is enzymatically removed with a decapping enzyme such as T4 polynucleotide kinase or tobacco acid pyrophosphatase. An oligonucleotide, which may be either a DNA oligonucleotide or a DNA-RNA hybrid oligonucleotide having RNA at its 3' end, is then ligated to the phosphate present at the 5' end of the decapped mRNA using T4 RNA ligase. The oligonucleotide may include a
15 restriction site to facilitate cloning of the cDNAs following their synthesis. Example 12 below describes one enzymatic method based on the doctoral thesis of Dumas.

EXAMPLE 12

Enzymatic Approach for Obtaining 5' ESTs

20 Twenty micrograms of PolyA+ RNA were dephosphorylated using Calf Intestinal Phosphatase (Biolabs). After a phenol chloroform extraction, the cap structure of mRNA was hydrolysed using the Tobacco Acid Pyrophosphatase (purified as described by Shinshi *et al.*, *Biochemistry* 15: 2185-2190, 1976) and a hemi 5'DNA/RNA-3' oligonucleotide having an unphosphorylated 5' end, a stretch of adenosine ribophosphate at the 3' end, and an EcoRI
25 site near the 5' end was ligated to the 5'P ends of mRNA using the T4 RNA ligase (Biolabs). Oligonucleotides suitable for use in this procedure are preferably 30 to 50 bases in length. Oligonucleotides having an unphosphorylated 5' end may be synthesized by adding a fluorochrome at the 5' end. The inclusion of a stretch of adenosine ribophosphates at the 3' end of the oligonucleotide increases ligation efficiency. It will be appreciated that the
30 oligonucleotide may contain cloning sites other than EcoRI.

Following ligation of the oligonucleotide to the phosphate present at the 5' end of the decapped mRNA, first and second strand cDNA synthesis is carried out using conventional methods or those specified in EP0 625,572 and Kato *et al. supra*, and Dumas Milne Edwards, *supra*, the disclosures of which are incorporated herein by reference. The resulting cDNA may then be ligated into vectors such as those disclosed in Kato *et al. supra* or other nucleic acid vectors known to those skilled in the art using techniques such as those described in Sambrook *et al.*, Molecular Cloning: A Laboratory Manual 2d Ed., Cold Spring Harbor Laboratory Press, 1989, the disclosure of which is incorporated herein by reference.

II. Obtention and Characterization of the 5' ESTs of the Present Invention

The 5' ESTs of the present invention were obtained using the aforementioned chemical and enzymatic approaches for enriching mRNAs for those having intact 5' ends as described below.

1. Obtention of 5' ESTS Using mRNAs with Intact 5' Ends

First, mRNAs were prepared as described in Example 13 below.

EXAMPLE 13

Preparation of mRNA With Intact 5' Ends

Total human RNAs or polyA⁺ RNAs derived from 29 different tissues were respectively purchased from LABIMO and CLONTECH and used to generate 44 cDNA libraries as follows. The purchased RNA had been isolated from cells or tissues using acid guanidium thiocyanate-phenol-chloroform extraction (Chomczynski and Sacchi, *Analytical Biochemistry* 162:156-159, 1987). PolyA⁺ RNA was isolated from total RNA (LABIMO) by two passes of oligo dT chromatography, as described by Aviv and Leder, *Proc. Natl. Acad. Sci. USA* 69:1408-1412, 1972 in order to eliminate ribosomal RNA.

The quality and the integrity of the polyA⁺ RNAs were checked. Northern blots hybridized with a globin probe were used to confirm that the mRNAs were not degraded. Contamination of the polyA⁺ mRNAs by ribosomal sequences was checked using Northern blots and a probe derived from the sequence of the 28S rRNA. Preparations of mRNAs with

less than 5% of rRNAs were used in library construction. To avoid constructing libraries with RNAs contaminated by exogenous sequences (prokaryotic or fungal), the presence of bacterial 16S ribosomal sequences or of two highly expressed fungal mRNAs was examined using PCR.

5 Following preparation of the mRNAs, the above described chemical and/or the enzymatic procedures for enriching mRNAs for those having intact 5' ends were employed to obtain 5' ESTs from various tissues. In both approaches, an oligonucleotide tag was attached to the 5' ends of the mRNAs. The oligonucleotide tag had an EcoRI site therein to facilitate later cloning procedures. To facilitate the processing of single stranded and double
10 stranded cDNA obtained in the construction of the libraries, the same nucleotidic sequence was used to design the ligated oligonucleotide in both chemical and enzymatic approaches. Nevertheless, in the chemical procedure, the tag used was an oligodeoxyribonucleotide which was linked to the cap of the mRNA whereas in the enzymatic ligation, the tag was a chimeric
15 hemi 5'DNA/RNA3' oligonucleotide which was ligated to the 5' end of decapped mRNA as described in example 12.

 Following attachment of the oligonucleotide tag to the mRNA by either the chemical or enzymatic methods, the integrity of the mRNA was examined by performing a Northern blot with 200 to 500 ng of mRNA using a probe complementary to the oligonucleotide tag before performing the first strand synthesis as described in example 14.

20

EXAMPLE 14

cDNA Synthesis Using mRNA Templates Having Intact 5' Ends

 For the mRNAs joined to oligonucleotide tags using both the chemical and enzymatic methods, first strand cDNA synthesis was performed using the Superscript II (Gibco BRL) or
25 the Rnase H Minus M-MLV (Promega) reverse transcriptase with random nonamers as primers. In order to protect internal EcoRI sites in the cDNA from digestion at later steps in the procedure, methylated dCTP was used for first strand synthesis. After removal of RNA by an alkaline hydrolysis, the first strand of cDNA was precipitated using isopropanol in order to eliminate residual primers.

30 For both the chemical and the enzymatic methods, the second strand of the cDNA was synthesized with a Klenow fragment using a primer corresponding to the 5' end of the

ligated oligonucleotide described in Example 12. Preferably, the primer is 20-25 bases in length. Methylated dCTP was also used for second strand synthesis in order to protect internal EcoRI sites in the cDNA from digestion during the cloning process.

Following cDNA synthesis, the cDNAs were cloned into pBlueScript as described in
5 Example 15 below.

EXAMPLE 15

Cloning of cDNAs derived from mRNA with intact 5' ends into BlueScript

Following second strand synthesis, the ends of the cDNA were blunted with T4 DNA
10 polymerase (Biolabs) and the cDNA was digested with EcoRI. Since methylated dCTP was used during cDNA synthesis, the EcoRI site present in the tag was the only hemi-methylated site, hence the only site susceptible to EcoRI digestion. The cDNA was then size fractionated using exclusion chromatography (AcA, Biosepra) and fractions corresponding to cDNAs of more than 150 bp were pooled and ethanol precipitated. The cDNA was directionally cloned
15 into the SmaI and EcoRI ends of the phagemid pBlueScript vector (Stratagene). The ligation mixture was electroporated into bacteria and propagated under appropriate antibiotic selection.

Clones containing the oligonucleotide tag attached were then selected as described in
Example 16 below.

20

EXAMPLE 16

Selection of Clones Having the Oligonucleotide Tag Attached Thereto

The plasmid DNAs containing 5' EST libraries made as described above were purified (Qiagen). A positive selection of the tagged clones was performed as follows.
25 Briefly, in this selection procedure, the plasmid DNA was converted to single stranded DNA using gene II endonuclease of the phage F1 in combination with an exonuclease (Chang *et al.*, *Gene* 127:95-8, 1993) such as exonuclease III or T7 gene 6 exonuclease. The resulting single stranded DNA was then purified using paramagnetic beads as described by Fry *et al.*, *Biotechniques*, 13: 124-131, 1992. In this procedure, the single stranded DNA was
30 hybridized with a biotinylated oligonucleotide having a sequence corresponding to the 3' end of the oligonucleotide described in Example 13. Preferably, the primer has a length of 20-25

bases. Clones including a sequence complementary to the biotinylated oligonucleotide were captured by incubation with streptavidin coated magnetic beads followed by magnetic selection. After capture of the positive clones, the plasmid DNA was released from the magnetic beads and converted into double stranded DNA using a DNA polymerase such as the ThermoSequenase obtained from Amersham Pharmacia Biotech. Alternatively, protocols such as the one described in the Gene Trapper kit available from Gibco BRL may be used. The double stranded DNA was then electroporated into bacteria. The percentage of positive clones having the 5' tag oligonucleotide was estimated to typically rank between 90 and 98% using dot blot analysis.

Following electroporation, the libraries were ordered in 384-microtiter plates (MTP). A copy of the MTP was stored for future needs. Then the libraries were transferred into 96 MTP and sequenced as described below.

EXAMPLE 17

Sequencing of Inserts in Selected Clones

Plasmid inserts were first amplified by PCR on PE 9600 thermocyclers (Perkin-Elmer, Applied Biosystems Division, Foster City, CA), using standard SETA-A and SETA-B primers (Genset SA), AmpliTaqGold (Perkin-Elmer), dNTPs (Boehringer), buffer and cycling conditions as recommended by the Perkin-Elmer Corporation.

PCR products were then sequenced using automatic ABI Prism 377 sequencers (Perkin Elmer). Sequencing reactions were performed using PE 9600 thermocyclers with standard dye-primer chemistry and ThermoSequenase (Amersham Pharmacia Biotech). The primers used were either T7 or 21M13 (available from Genset SA) as appropriate. The primers were labeled with the JOE, FAM, ROX and TAMRA dyes. The dNTPs and ddNTPs used in the sequencing reactions were purchased from Boehringer. Sequencing buffer, reagent concentrations and cycling conditions were as recommended by Amersham.

Following the sequencing reaction, the samples were precipitated with ethanol, resuspended in formamide loading buffer, and loaded on a standard 4% acrylamide gel. Electrophoresis was performed for 2.5 hours at 3000V on an ABI 377 sequencer, and the sequence data were collected and analyzed using the ABI Prism DNA Sequencing Analysis Software, version 2.1.2.

2. Computer analysis of the Obtained 5' ESTs: Construction of NetGene and SignalTag databases

The sequence data from the 44 cDNA libraries made as described above were transferred to a proprietary database, where quality control and validation steps were performed. A proprietary base-caller, working using a Unix system, automatically flagged suspect peaks, taking into account the shape of the peaks, the inter-peak resolution, and the noise level. The proprietary base-caller also performed an automatic trimming. Any stretch of 25 or fewer bases having more than 4 suspect peaks was considered unreliable and was discarded. Sequences corresponding to cloning vector or ligation oligonucleotides were automatically removed from the EST sequences. However, the resulting EST sequences may contain 1 to 5 bases belonging to the above mentioned sequences at their 5' end. If needed, these can easily be removed on a case to case basis.

Following sequencing as described above, the sequences of the 5' ESTs were entered in NetGene™, a proprietary database called for storage and manipulation as described below. It will be appreciated by those skilled in the art that the data could be stored and manipulated on any medium which can be read and accessed by a computer. Computer readable media include magnetically, optically, or electronically readable media. For example, the computer readable media may be a hard disc, a floppy disc, a magnetic tape, CD-ROM, RAM, or ROM as well as other types of other media known to those skilled in the art.

In addition, the sequence data may be stored and manipulated in a variety of data processor programs in a diversity of formats. For instance, the sequence data may be stored as text in a word processing file, such as Microsoft WORD or WORDPERFECT or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2, SYBASE, or ORACLE.

The computer readable media on which the sequence information is stored may be in a personal computer, a network, a server or other computer systems known to those skilled in the art. The computer or other system preferably includes the storage media described above, and a processor for accessing and manipulating the sequence data. Once the sequence data has been stored, it may be manipulated and searched to locate those stored sequences which contain a desired nucleic acid sequence or which encode a protein having a particular functional domain. For example, the stored sequence information may be compared to other

known sequences to identify homologies, motifs implicated in biological function, or structural motifs.

Programs which may be used to search or compare the stored sequences include the MacPattern (EMBL), BLAST, and BLAST2 program series (NCBI), basic local alignment
5 search tool programs for nucleotide (BLASTN) and peptide (BLASTX) comparisons (Altschul *et al*, *J. Mol. Biol.* 215: 403, 1990) and FASTA (Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85: 2444, 1988). The BLAST programs then extend the alignments on the basis of defined match and mismatch criteria.

Motifs which may be detected using the above programs and those described in
10 Example 28 include sequences encoding leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

15 Before searching the cDNAs in the NetGene™ database for sequence motifs of interest, cDNAs derived from mRNAs which were not of interest were identified and eliminated from further consideration as described in Example 18 below.

EXAMPLE 18

20 Elimination of Undesired Sequences from Further Consideration

5' ESTs in the NetGene™ database which were derived from undesired sequences such as transfer RNAs, ribosomal RNAs, mitochondrial RNAs, prokaryotic RNAs, fungal RNAs, Alu sequences, L1 sequences, or repeat sequences were identified using the FASTA and BLASTN programs with the parameters listed in Table I.

25 To eliminate 5' ESTs encoding tRNAs from further consideration, the 5' EST sequences were compared to the sequences of 1190 known tRNAs obtained from EMBL release 38, of which 100 were human. The comparison was performed using FASTA on both strands of the 5' ESTs. Sequences having more than 80% homology over more than 60 nucleotides were identified as tRNA. Of the 144,341 sequences screened, 26 were identified
30 as tRNAs and eliminated from further consideration.

To eliminate 5' ESTs encoding rRNAs from further consideration, the 5' EST sequences were compared to the sequences of 2497 known rRNAs obtained from EMBL release 38, of which 73 were human. The comparison was performed using BLASTN on both strands of the 5' ESTs with the parameter S=108. Sequences having more than 80%
5 homology over stretches longer than 40 nucleotides were identified as rRNAs. Of the 144,341 sequences screened, 3,312 were identified as rRNAs and eliminated from further consideration.

To eliminate 5' ESTs encoding mtRNAs from further consideration, the 5' EST sequences were compared to the sequences of the two known mitochondrial genomes for
10 which the entire genomic sequences are available and all sequences transcribed from these mitochondrial genomes including tRNAs, rRNAs, and mRNAs for a total of 38 sequences. The comparison was performed using BLASTN on both strands of the 5' ESTs with the parameter S=108. Sequences having more than 80% homology over stretches longer than 40 nucleotides were identified as mtRNAs. Of the 144,341 sequences screened, 6,110 were
15 identified as mtRNAs and eliminated from further consideration.

Sequences which might have resulted from exogenous contaminants were eliminated from further consideration by comparing the 5' EST sequences to release 46 of the EMBL bacterial and fungal divisions using BLASTN with the parameter S=144. All sequences having more than 90% homology over at least 40 nucleotides were identified as exogenous
20 contaminants. Of the 42 cDNA libraries examined, the average percentages of prokaryotic and fungal sequences contained therein were 0.2% and 0.5% respectively. Among these sequences, only one could be identified as a sequence specific to fungi. The others were either fungal or prokaryotic sequences having homologies with vertebrate sequences or including repeat sequences which had not been masked during the electronic comparison.

25 In addition, the 5' ESTs were compared to 6093 Alu sequences and 1115 L1 sequences to mask 5' ESTs containing such repeat sequences. 5' ESTs including THE and MER repeats, SSTR sequences or satellite, micro-satellite, or telomeric repeats were also eliminated from further consideration. On average, 11.5% of the sequences in the libraries contained repeat sequences. Of this 11.5%, 7% contained Alu repeats, 3.3% contained L1
30 repeats and the remaining 1.2% were derived from the other screened types of repetitive sequences. These percentages are consistent with those found in cDNA libraries prepared by

other groups. For example, the cDNA libraries of Adams *et al.* contained between 0% and 7.4% Alu repeats depending on the source of the RNA which was used to prepare the cDNA library (Adams *et al.*, *Nature* 377:174, 1996).

- 5 The sequences of those 5' ESTs remaining after the elimination of undesirable sequences were compared with the sequences of known human mRNAs to determine the accuracy of the sequencing procedures described above.

EXAMPLE 19

10 Measurement of Sequencing Accuracy by Comparison to Known Sequences

To further determine the accuracy of the sequencing procedure described above, the sequences of 5' ESTs derived from known sequences were identified and compared to the original known sequences. First, a FASTA analysis with overhangs shorter than 5 bp on both ends was conducted on the 5' ESTs to identify those matching an entry in the public human mRNA database. The 6655 5' ESTs which matched a known human mRNA were then realigned with their cognate mRNA and dynamic programming was used to include substitutions, insertions, and deletions in the list of "errors" which would be recognized. Errors occurring in the last 10 bases of the 5' EST sequences were ignored to avoid the inclusion of spurious cloning sites in the analysis of sequencing accuracy.

- 20 This analysis revealed that the sequences incorporated in the NetGene™ database had an accuracy of more than 99.5%.

To determine the efficiency with which the above selection procedures select cDNAs which include the 5' ends of their corresponding mRNAs, the following analysis was performed.

25

EXAMPLE 20

Determination of Efficiency of 5' EST Selection

- To determine the efficiency at which the above selection procedures isolated 5' ESTs which included sequences close to the 5' end of the mRNAs from which they derived, the sequences of the ends of the 5' ESTs derived from the elongation factor 1 subunit α and
- 30

ferritin heavy chain genes were compared to the known cDNA sequences of these genes. Since the transcription start sites of both genes are well characterized, they may be used to determine the percentage of derived 5' ESTs which included the authentic transcription start sites.

5 For both genes, more than 95% of the obtained 5' ESTs actually included sequences close to or upstream of the 5' end of the corresponding mRNAs.

To extend the analysis of the reliability of the procedures for isolating 5' ESTs from ESTs in the NetGene™ database, a similar analysis was conducted using a database composed of human mRNA sequences extracted from GenBank database release 97 for
10 comparison. The 5' ends of more than 85% of 5' ESTs derived from mRNAs included in the GeneBank database were located close to the 5' ends of the known sequence. As some of the mRNA sequences available in the GenBank database are deduced from genomic sequences, a 5' end matching with these sequences will be counted as an internal match. Thus, the method used here underestimates the yield of ESTs including the authentic 5' ends
15 of their corresponding mRNAs.

The EST libraries made above included multiple 5' ESTs derived from the same mRNA. The sequences of such 5' ESTs were compared to one another and the longest 5' ESTs for each mRNA were identified. Overlapping cDNAs were assembled into continuous
20 sequences (contigs). The resulting continuous sequences were then compared to public databases to gauge their similarity to known sequences, as described in Example 21 below.

EXAMPLE 21

Clustering of the 5' ESTs and Calculation of Novelty Indices for cDNA Libraries

25 For each sequenced EST library, the sequences were clustered by the 5' end. Each sequence in the library was compared to the others with BLASTN2 (direct strand, parameters S=107). ESTs with High Scoring Segment Pairs (HSPs) at least 25 bp long, having 95% identical bases and beginning closer than 10 bp from each EST 5' end were grouped. The longest sequence found in the cluster was used as representative of the group. A global
30 clustering between libraries was then performed leading to the definition of super-contigs.

To assess the yield of new sequences within the EST libraries, a novelty rate (NR) was defined as: $NR = 100 \times (\text{Number of new unique sequences found in the library} / \text{Total number of sequences from the library})$. Typically, novelty rating ranged between 10% and 41% depending on the tissue from which the EST library was obtained. For most of the
5 libraries, the random sequencing of 5' EST libraries was pursued until the novelty rate reached 20%.

Following characterization as described above, the collection of 5' ESTs in NetGene™ was screened to identify those 5' ESTs bearing potential signal sequences as
10 described in Example 22 below.

EXAMPLE 22

Identification of Potential Signal Sequences in 5' ESTs

The 5' ESTs in the NetGene™ database were screened to identify those having an
15 uninterrupted open reading frame (ORF) longer than 45 nucleotides beginning with an ATG codon and extending to the end of the EST. Approximately half of the cDNA sequences in NetGene™ contained such an ORF. The ORFs of these 5' ESTs were then searched to identify potential signal motifs using slight modifications of the procedures disclosed in Von Heijne, *Nucleic Acids Res.* 14:4683-4690, 1986, the disclosure of which is incorporated
20 herein by reference. Those 5' EST sequences encoding a stretch of at least 15 amino acid long with a score of at least 3.5 in the Von Heijne signal peptide identification matrix were considered to possess a signal sequence. Those 5' ESTs which matched a known human mRNA or EST sequence and had a 5' end more than 20 nucleotides downstream of the known 5' end were excluded from further analysis. The remaining cDNAs having signal
25 sequences therein were included in a database called SignalTag™.

To confirm the accuracy of the above method for identifying signal sequences, the analysis of Example 23 was performed.

EXAMPLE 23**Confirmation of Accuracy of Identification of Potential Signal Sequences in 5' ESTs**

The accuracy of the above procedure for identifying signal sequences encoding signal peptides was evaluated by applying the method to the 43 amino acids located at the N terminus of all human SwissProt proteins. The computed Von Heijne score for each protein was compared with the known characterization of the protein as being a secreted protein or a non-secreted protein. In this manner, the number of non-secreted proteins having a score higher than 3.5 (false positives) and the number of secreted proteins having a score lower than 3.5 (false negatives) could be calculated.

Using the results of the above analysis, the probability that a peptide encoded by the 5' region of the mRNA is in fact a genuine signal peptide based on its Von Heijne's score was calculated based on either the assumption that 10 % of human proteins are secreted or the assumption that 20 % of human proteins are secreted. The results of this analysis are shown in Figure 2 and in table IV.

Using the above method of identification of secretory proteins, 5' ESTs of the following polypeptides known to be secreted were obtained: human glucagon, gamma interferon induced monokine precursor, secreted cyclophilin-like protein, human pleiotropin, and human biotinidase precursor. Thus, the above method successfully identified those 5' ESTs which encode a signal peptide.

To confirm that the signal peptide encoded by the 5' ESTs actually functions as a signal peptide, the signal sequences from the 5' ESTs may be cloned into a vector designed for the identification of signal peptides. Such vectors are designed to confer the ability to grow in selective medium only to host cells containing a vector with an operably linked signal sequence. For example, to confirm that a 5' EST encodes a genuine signal peptide, the signal sequence of the 5' EST may be inserted upstream and in frame with a non-secreted form of the yeast invertase gene in signal peptide selection vectors such as those described in U.S. Patent No. 5,536,637, the disclosure of which is incorporated herein by reference. Growth of host cells containing signal sequence selection vectors with the correctly inserted 5' EST signal sequence confirms that the 5' EST encodes a genuine signal peptide.

Alternatively, the presence of a signal peptide may be confirmed by cloning the extended cDNAs obtained using the ESTs into expression vectors such as pXT1 (as described below in example 30), or by constructing promoter-signal sequence-reporter gene vectors which encode fusion proteins between the signal peptide and an assayable reporter protein. After introduction of these vectors into a suitable host cell, such as COS cells or NIH 3T3 cells, the growth medium may be harvested and analyzed for the presence of the secreted protein. The medium from these cells is compared to the medium from control cells containing vectors lacking the signal sequence or extended cDNA insert to identify vectors which encode a functional signal peptide or an authentic secreted protein.

Those 5' ESTs which encoded a signal peptide, as determined by the method of Example 22 above, were further grouped into four categories based on their homology to known sequences as described in Example 24 below.

EXAMPLE 24

Categorization of 5' ESTs Encoding a Signal Peptide

Those 5' ESTs having a sequence not matching any known vertebrate sequence nor any publicly available EST sequence were designated "new." Of the sequences in the SignalTag™ database, 947 of the 5' ESTs having a Von Heijne's score of at least 3.5 fell into this category.

Those 5' ESTs having a sequence not matching any vertebrate sequence but matching a publicly known EST were designated "EST-ext", provided that the known EST sequence was extended by at least 40 nucleotides in the 5' direction. Of the sequences in the SignalTag™ database, 150 of the 5' ESTs having a Von Heijne's score of at least 3.5 fell into this category.

Those ESTs not matching any vertebrate sequence but matching a publicly known EST without extending the known EST by at least 40 nucleotides in the 5' direction were designated "EST." Of the sequences in the SignalTag™ database, 599 of the 5' ESTs having a Von Heijne's score of at least 3.5 fell into this category.

Those 5' ESTs matching a human mRNA sequence but extending the known sequence by at least 40 nucleotides in the 5' direction were designated "VERT-ext." Of the sequences in the SignalTag™ database, 23 of the 5' ESTs having a Von Heijne's score of at

least 3.5 fell into this category. Included in this category was a 5' EST which extended the known sequence of the human translocase mRNA by more than 200 bases in the 5' direction. A 5' EST which extended the sequence of a human tumor suppressor gene in the 5' direction was also identified.

5 Table V shows the distribution of 5' ESTs in each category and the number of 5' ESTs in each category having a given minimum von Heijne's score.

3. Evaluation of Spatial and Temporal Expression of mRNAs Corresponding to the 5'ESTs or Extended cDNAs

10

Each of the 5' ESTs was also categorized based on the tissue from which its corresponding mRNA was obtained, as described below in Example 25.

EXAMPLE 25

15

Categorization of Expression Patterns

Table VI shows the distribution of 5' ESTs in each of the above defined category with respect to the tissue from which the 5'ESTs of the corresponding mRNA were obtained.

20 Table II provides the sequence identification numbers of 5' EST sequences derived from different tissues, the categories in which these sequences fall, and the von Heijne's score of the signal peptides which they encode. The 5' EST sequences and the amino acid sequences they encode are provided in the appended sequence listings. Table III provides the sequence ID numbers of the 5' ESTs and the sequences of the signal peptides which they encode. The sequences of the 5' ESTs and the polypeptides they encode are provided in the sequence listing appended hereto.

25 The sequences of DNA SEQ ID NOs: 38-291 can readily be screened for any errors therein and any sequence ambiguities can be resolved by resequencing a fragment containing such errors or ambiguities on both strands. Such fragments may be obtained from the plasmids stored in the inventors' laboratory or can be isolated using the techniques described herein. Resolution of any such ambiguities or errors may be facilitated by using primers
30 which hybridize to sequences located close to the ambiguous or erroneous sequences. For example, the primers may hybridize to sequences within 50-75 bases of the ambiguity or

error. Upon resolution of an error or ambiguity, the corresponding corrections can be made in the protein sequences encoded by the DNA containing the error or ambiguity.

In addition to categorizing the 5' ESTs with respect to their tissue of origin, the spatial and temporal expression patterns of the mRNAs corresponding to the 5' ESTs, as well as their expression levels, may be determined as described in Example 26 below. Characterization of the spatial and temporal expression patterns and expression levels of these mRNAs is useful for constructing expression vectors capable of producing a desired level of gene product in a desired spatial or temporal manner, as will be discussed in more detail below.

Furthermore, 5' ESTs whose corresponding mRNAs are associated with disease states may also be identified. For example, a particular disease may result from the lack of expression, over expression, or under expression of an mRNA corresponding to a 5' EST. By comparing mRNA expression patterns and quantities in samples taken from healthy individuals with those from individuals suffering from a particular disease, 5' ESTs responsible for the disease may be identified.

It will be appreciated that the results of the above characterization procedures for 5' ESTs also apply to extended cDNAs (obtainable as described below) which contain sequences adjacent to the 5' ESTs. It will also be appreciated that if desired, characterization may be delayed until extended cDNAs have been obtained rather than characterizing the ESTs themselves.

EXAMPLE 26

Evaluation of Expression Levels and Patterns of mRNAs

Corresponding to 5' ESTs or Extended cDNAs

Expression levels and patterns of mRNAs corresponding to 5' ESTs or extended cDNAs (obtainable as described below in example 27) may be analyzed by solution hybridization with long probes as described in International Patent Application No. WO 97/05277, the entire contents of which are hereby incorporated by reference. Briefly, a 5' EST, extended cDNA, or fragment thereof corresponding to the gene encoding the mRNA to be characterized is inserted at a cloning site immediately downstream of a bacteriophage (T3,

T7 or SP6) RNA polymerase promoter to produce antisense RNA. Preferably, the 5' EST or extended cDNA has 100 or more nucleotides. The plasmid is linearized and transcribed in the presence of ribonucleotides comprising modified ribonucleotides (*i.e.* biotin-UTP and DIG-UTP). An excess of this doubly labeled RNA is hybridized in solution with mRNA isolated
5 from cells or tissues of interest. The hybridizations are performed under standard stringent conditions (40-50°C for 16 hours in an 80% formamide, 0.4 M NaCl buffer, pH 7-8). The unhybridized probe is removed by digestion with ribonucleases specific for single-stranded RNA (*i.e.* RNases CL3, T1, Phy M, U2 or A). The presence of the biotin-UTP modification enables capture of the hybrid on a microtitration plate coated with streptavidin. The presence
10 of the DIG modification enables the hybrid to be detected and quantified by ELISA using an anti-DIG antibody coupled to alkaline phosphatase.

The 5' ESTs, extended cDNAs, or fragments thereof may also be tagged with nucleotide sequences for the serial analysis of gene expression (SAGE) as disclosed in UK Patent Application No. 2 305 241 A, the entire contents of which are incorporated by
15 reference. In this method, cDNAs are prepared from a cell, tissue, organism or other source of nucleic acid for which gene expression patterns must be determined. The resulting cDNAs are separated into two pools. The cDNAs in each pool are cleaved with a first restriction endonuclease, called an anchoring enzyme, having a recognition site which is likely to be present at least once in most cDNAs. The fragments which contain the 5' or 3' most region
20 of the cleaved cDNA are isolated by binding to a capture medium such as streptavidin coated beads. A first oligonucleotide linker having a first sequence for hybridization of an amplification primer and an internal restriction site for a so-called tagging endonuclease is ligated to the digested cDNAs in the first pool. Digestion with the second endonuclease produces short tag fragments from the cDNAs.

25 A second oligonucleotide having a second sequence for hybridization of an amplification primer and an internal restriction site is ligated to the digested cDNAs in the second pool. The cDNA fragments in the second pool are also digested with the tagging endonuclease to generate short tag fragments derived from the cDNAs in the second pool. The tags resulting from digestion of the first and second pools with the anchoring enzyme and
30 the tagging endonuclease are ligated to one another to produce so-called ditags. In some embodiments, the ditags are concatamerized to produce ligation products containing from 2

to 200 ditags. The tag sequences are then determined and compared to the sequences of the 5' ESTs or extended cDNAs to determine which 5' ESTs or extended cDNAs are expressed in the cell, tissue, organism, or other source of nucleic acids from which the tags were derived. In this way, the expression pattern of the 5' ESTs or extended cDNAs in the cell,
5 tissue, organism, or other source of nucleic acids is obtained.

Quantitative analysis of gene expression may also be performed using arrays. As used herein, the term array means a one dimensional, two dimensional, or multidimensional arrangement of full length cDNAs (*i.e.* extended cDNAs which include the coding sequence for the signal peptide, the coding sequence for the mature protein, and a stop codon),
10 extended cDNAs, 5' ESTs or fragments thereof of sufficient length to permit specific detection of gene expression. Preferably, the fragments are at least 15 nucleotides in length. More preferably, the fragments are at least 100 nucleotide long. More preferably, the fragments are more than 100 nucleotides in length. In some embodiments, the fragments may be more than 500 nucleotide long.

15 For example, quantitative analysis of gene expression may be performed with full length cDNAs as defined below, extended cDNAs, 5' ESTs, or fragments thereof in a complementary DNA microarray as described by Schena *et al.* (*Science* 270:467-470, 1995; *Proc. Natl. Acad. Sci. U.S.A.* 93:10614-10619, 1996). Full length cDNAs, extended cDNAs, 5' ESTs or fragments thereof are amplified by PCR and arrayed from 96-well microtiter
20 plates onto silylated microscope slides using high-speed robotics. Printed arrays are incubated in a humid chamber to allow rehydration of the array elements and rinsed, once in 0.2% SDS for 1 min, twice in water for 1 min and once for 5 min in sodium borohydride solution. The arrays are submerged in water for 2 min at 95°C, transferred into 0.2% SDS for 1 min, rinsed twice with water, air dried and stored in the dark at 25°C.

25 Cell or tissue mRNA is isolated or commercially obtained and probes are prepared by a single round of reverse transcription. Probes are hybridized to 1 cm² microarrays under a 14 x 14 mm glass coverslip for 6-12 hours at 60°C. Arrays are washed for 5 min at 25°C in low stringency wash buffer (1 x SSC/0.2% SDS), then for 10 min at room temperature in high stringency wash buffer (0.1 x SSC/0.2% SDS). Arrays are scanned in 0.1 x SSC using a
30 fluorescence laser scanning device fitted with a custom filter set. Accurate differential

expression measurements are obtained by taking the average of the ratios of two independent hybridizations.

Quantitative analysis of the expression of genes may also be performed with full length cDNAs, extended cDNAs, 5' ESTs, or fragments thereof in complementary DNA arrays as described by Pietu *et al.* (*Genome Research* 6:492-503, 1996). The full length
5 cDNAs, extended cDNAs, 5' ESTs or fragments thereof are PCR amplified and spotted on membranes. Then, mRNAs originating from various tissues or cells are labeled with radioactive nucleotides. After hybridization and washing in controlled conditions, the hybridized mRNAs are detected by phospho-imaging or autoradiography. Duplicate
10 experiments are performed and a quantitative analysis of differentially expressed mRNAs is then performed.

Alternatively, expression analysis of the 5' ESTs or extended cDNAs can be done through high density nucleotide arrays as described by Lockhart *et al.* (*Nature Biotechnology* 14: 1675-1680, 1996) and Sosnowsky *et al.* (*Proc. Natl. Acad. Sci.* 94:1119-1123, 1997).
15 Oligonucleotides of 15-50 nucleotides corresponding to sequences of the 5' ESTs or extended cDNAs are synthesized directly on the chip (Lockhart *et al.*, *supra*) or synthesized and then addressed to the chip (Sosnowsky *et al.*, *supra*). Preferably, the oligonucleotides are about 20 nucleotides in length.

cDNA probes labeled with an appropriate compound, such as biotin, digoxigenin
20 or fluorescent dye, are synthesized from the appropriate mRNA population and then randomly fragmented to an average size of 50 to 100 nucleotides. The said probes are then hybridized to the chip. After washing as described in Lockhart *et al.*, *supra* and application of different electric fields (Sonowsky *et al.*, *supra.*), the dyes or labeling compounds are detected and quantified. Duplicate hybridizations are performed.
25 Comparative analysis of the intensity of the signal originating from cDNA probes on the same target oligonucleotide in different cDNA samples indicates a differential expression of the mRNA corresponding to the 5' EST or extended cDNA from which the oligonucleotide sequence has been designed.

III. Use of 5' ESTs to Clone Extended cDNAs and to Clone the Corresponding Genomic DNAs

Once 5' ESTs which include the 5' end of the corresponding mRNAs have been selected using the procedures described above, they can be utilized to isolate extended cDNAs which contain sequences adjacent to the 5' ESTs. The extended cDNAs may include the entire coding sequence of the protein encoded by the corresponding mRNA, including the authentic translation start site, the signal sequence, and the sequence encoding the mature protein remaining after cleavage of the signal peptide. Such extended cDNAs are referred to herein as "full length cDNAs." Alternatively, the extended cDNAs may include only the sequence encoding the mature protein remaining after cleavage of the signal peptide, or only the sequence encoding the signal peptide.

Example 27 below describes a general method for obtaining extended cDNAs using 5' ESTs. Example 28 below provides experimental results, using the method explained in example 27, describing several extended cDNAs including the entire coding sequence and authentic 5' end of the corresponding mRNA for several secreted proteins.

The methods of Examples 27, 28, and 29 can also be used to obtain extended cDNAs which encode less than the entire coding sequence of the secreted proteins encoded by the genes corresponding to the 5' ESTs. In some embodiments, the extended cDNAs isolated using these methods encode at least 10 amino acids of one of the proteins encoded by the sequences of SEQ ID NOs: 38-291. In further embodiments, the extended cDNAs encode at least 20 amino acids of the proteins encoded by the sequences of SEQ ID NOs: 38-291. In further embodiments, the extended cDNAs encode at least 30 amino acids of the sequences of SEQ ID NOs: 38-291. In a preferred embodiment, the extended cDNAs encode a full length protein sequence, which includes the protein coding sequences of SEQ ID NOs: 38-291.

EXAMPLE 27

General Method for Using 5' ESTs to Clone and Sequence cDNAs which Include the Entire Coding Region and the Authentic 5' End of the Corresponding mRNA

The following general method has been used to quickly and efficiently isolate extended cDNAs having the authentic 5' ends of their corresponding mRNAs as well as

the full protein coding sequence and including sequence adjacent to the sequences of the 5' ESTs used to obtain them. This method may be applied to obtain extended cDNAs for any 5' EST in the NetGene™ database, including those 5' ESTs encoding polypeptides belonging to secreted proteins. The method is summarized in figure 3.

5

1. Obtention of Extended cDNAs

a) First strand synthesis

The method takes advantage of the known 5' sequence of the mRNA. A reverse transcription reaction is conducted on purified mRNA with a poly 14dT primer containing a 49 nucleotide sequence at its 5' end allowing the addition of a known sequence at the end of the cDNA which corresponds to the 3' end of the mRNA. For example, the primer may have the following sequence: 5'-ATC GTT GAG ACT CGT ACC AGC AGA GTC ACG AGA GAG ACT ACA CGG TAC TGG TTT TTT TTT TTVN -3' (SEQ ID NO:14). Those skilled in the art will appreciate that other sequences may also be added to the poly dT sequence and used to prime the first strand synthesis. Using this primer and a reverse transcriptase such as the Superscript II (Gibco BRL) or Rnase H Minus M-MLV (Promega) enzyme, a reverse transcript anchored at the 3' polyA site of the RNAs is generated.

After removal of the mRNA hybridized to the first cDNA strand by alkaline hydrolysis, the products of the alkaline hydrolysis and the residual poly dT primer are eliminated with an exclusion column such as an AcA34 (Biosepra) matrix as explained in Example 11.

b) Second strand synthesis

A pair of nested primers on each end is designed based on the known 5' sequence from the 5' EST and the known 3' end added by the poly dT primer used in the first strand synthesis. Softwares used to design primers are either based on GC content and melting temperatures of oligonucleotides, such as OSP (Illier and Green, *PCR Meth. Appl.* 1:124-128, 1991), or based on the octamer frequency disparity method (Griffais *et al.*, *Nucleic Acids Res.* 19: 3887-3891, 1991) such as PC-Rare (<http://bioinformatics.weizmann.ac.il/software/PC-Rare/doc/manuel.html>).

Preferably, the nested primers at the 5' end are separated from one another by four to nine bases. The 5' primer sequences may be selected to have melting temperatures and specificities suitable for use in PCR.

Preferably, the nested primers at the 3' end are separated from one another by four to nine bases. For example, the nested 3' primers may have the following sequences: (5'- CCA GCA GAG TCA CGA GAG AGA CTA CAC GG -3'(SEQ ID NO:15), and 5'- CAC GAG AGA GAC TAC ACG GTA CTG G -3' (SEQ ID NO:16). These primers were selected because they have melting temperatures and specificities compatible with their use in PCR. However, those skilled in the art will appreciate that other sequences may also be used as primers.

The first PCR run of 25 cycles is performed using the Advantage Tth Polymerase Mix (Clontech) and the outer primer from each of the nested pairs. A second 20 cycle PCR using the same enzyme and the inner primer from each of the nested pairs is then performed on 1/2500 of the first PCR product. Thereafter, the primers and nucleotides are removed.

2. Sequencing of Full Length Extended cDNAs or Fragments Thereof

Due to the lack of position constraints on the design of 5' nested primers compatible for PCR use using the OSP software, amplicons of two types are obtained. Preferably, the second 5' primer is located upstream of the translation initiation codon thus yielding a nested PCR product containing the whole coding sequence. Such a full length extended cDNA undergoes a direct cloning procedure as described in section a. However, in some cases, the second 5' primer is located downstream of the translation initiation codon, thereby yielding a PCR product containing only part of the ORF. Such incomplete PCR products are submitted to a modified procedure described in section b.

a) Nested PCR products containing complete ORFs

When the resulting nested PCR product contains the complete coding sequence, as predicted from the 5'EST sequence, it is cloned in an appropriate vector such as pED6dpc2, as described in section 3.

b) Nested PCR products containing incomplete ORFs

When the amplicon does not contain the complete coding sequence, intermediate steps are necessary to obtain both the complete coding sequence and a PCR product

containing the full coding sequence. The complete coding sequence can be assembled from several partial sequences determined directly from different PCR products as described in the following section.

Once the full coding sequence has been completely determined, new primers
5 compatible for PCR use are designed to obtain amplicons containing the whole coding region. However, in such cases, 3' primers compatible for PCR use are located inside the 3' UTR of the corresponding mRNA, thus yielding amplicons which lack part of this region, *i.e.* the polyA tract and sometimes the polyadenylation signal, as illustrated in figure 3. Such full length extended cDNAs are then cloned into an appropriate vector as
10 described in section 3.

c) Sequencing extended cDNAs

Sequencing of extended cDNAs is performed using a Die Terminator approach with the AmpliTaq DNA polymerase FS kit available from Perkin Elmer.

In order to sequence PCR fragments, primer walking is performed using software
15 such as OSP to choose primers and automated computer software such as ASMG (Sutton *et al.*, *Genome Science Technol.* 1: 9-19, 1995) to construct contigs of walking sequences including the initial 5' tag using minimum overlaps of 32 nucleotides. Preferably, primer walking is performed until the sequences of full length cDNAs are obtained.

Completion of the sequencing of a given extended cDNA fragment is assessed as
20 follows. Since sequences located after a polyA tract are difficult to determine precisely in the case of uncloned products, sequencing and primer walking processes for PCR products are interrupted when a polyA tract is identified in extended cDNAs obtained as described in case b. The sequence length is compared to the size of the nested PCR product obtained as described above. Due to the limited accuracy of the determination of the PCR product size
25 by gel electrophoresis, a sequence is considered complete if the size of the obtained sequence is at least 70 % the size of the first nested PCR product. If the length of the sequence determined from the computer analysis is not at least 70 % of the length of the nested PCR product, these PCR products are cloned and the sequence of the insertion is determined. When Northern blot data are available, the size of the mRNA detected for a given PCR
30 product is used to finally assess that the sequence is complete. Sequences which do not fulfill the above criteria are discarded and will undergo a new isolation procedure.

Sequence data of all extended cDNAs are then transferred to a proprietary database, where quality controls and validation steps are carried out as described in example 15.

3. Cloning of Full Length Extended cDNAs

5 The PCR product containing the full coding sequence is then cloned in an appropriate vector. For example, the extended cDNAs can be cloned into the expression vector pED6dpc2 (DiscoverEase, Genetics Institute, Cambridge, MA) as follows. pED6dpc2 vector DNA is prepared with blunt ends by performing an EcoRI digestion followed by a fill
10 in reaction. The blunt ended vector is dephosphorylated. After removal of PCR primers and ethanol precipitation, the PCR product containing the full coding sequence or the extended cDNA obtained as described above is phosphorylated with a kinase subsequently removed by phenol-Sevag extraction and precipitation. The double stranded extended cDNA is then
ligated to the vector and the resulting expression plasmid introduced into appropriate host cells.

15 Since the PCR products obtained as described above are blunt ended molecules that can be cloned in either direction, the orientation of several clones for each PCR product is determined. Then, 4 to 10 clones are ordered in microtiter plates and subjected to a PCR reaction using a first primer located in the vector close to the cloning site and a second primer located in the portion of the extended cDNA corresponding to the 3' end of the mRNA. This
20 second primer may be the antisense primer used in anchored PCR in the case of direct cloning (case a) or the antisense primer located inside the 3'UTR in the case of indirect cloning (case b). Clones in which the start codon of the extended cDNA is operably linked to the promoter in the vector so as to permit expression of the protein encoded by the extended cDNA are conserved and sequenced. In addition to the ends of cDNA inserts, approximately 50 bp of
25 vector DNA on each side of the cDNA insert are also sequenced.

 The cloned PCR products are then entirely sequenced according to the aforementioned procedure. In this case, contiguation of long fragments is then performed on walking sequences that have already contiguated for uncloned PCR products during primer walking. Sequencing of cloned amplicons is complete when the resulting contigs
30 include the whole coding region as well as overlapping sequences with vector DNA on both ends.

4. Computer analysis of Full Length Extended cDNA

Sequences of all full length extended cDNAs are then submitted to further analysis as described below. Before searching the extended full length cDNAs for sequences of interest, extended cDNAs which are not of interest (vector RNAs, transfer RNAs, ribosomal RNAs, mitochondrial RNAs, prokaryotic RNAs and fungal RNAs) are discarded using methods essentially similar to those described for 5'ESTs in Example 18.

a) Identification of structural features

Structural features, e.g. polyA tail and polyadenylation signal, of the sequences of full length extended cDNAs are subsequently determined as follows.

10 A polyA tail is defined as a homopolymeric stretch of at least 11 A with at most one alternative base within it. The polyA tail search is restricted to the last 100 nt of the sequence and limited to stretches of 11 consecutive A's because sequencing reactions are often not readable after such a polyA stretch. Stretches having more than 90% homology over 8 nucleotides are identified as polyA tails using BLAST2N.

15 To search for a polyadenylation signal, the polyA tail is clipped from the full-length sequence. The 50 bp preceding the polyA tail are first searched for the canonic polyadenylation AAUAAA signal and, if the canonic signal is not detected, for the alternative AUUAAA signal (Sheets *et al.*, *Nuc. Acids Res.* 18: 5799-5805, 1990). If neither of these consensus polyadenylation signals is found, the canonic motif is searched again allowing one mismatch to account for possible sequencing errors. More than 85 % of identified polyadenylation signals of either type actually ends 10 to 30 bp from the polyA tail. Alternative AUUAAA signals represents approximately 15 % of the total number of identified polyadenylation signals.

b) Identification of functional features

25 Functional features, e.g. ORFs and signal sequences, of the sequences of full length extended cDNAs were subsequently determined as follows.

The 3 upper strand frames of extended cDNAs are searched for ORFs defined as the maximum length fragments beginning with a translation initiation codon and ending with a stop codon. ORFs encoding at least 20 amino acids are preferred.

30 Each found ORF is then scanned for the presence of a signal peptide in the first 50 amino-acids or, where appropriate, within shorter regions down to 20 amino acids or

less in the ORF, using the matrix method of von Heijne (*Nuc. Acids Res.* 14: 4683-4690, 1986), the disclosure of which is incorporated herein by reference as described in Example 22.

c) Homology to either nucleotidic or proteic sequences

- 5 Categorization of full-length sequences may be achieved using procedures essentially similar to those described for 5'ESTs in Example 24.

Extended cDNAs prepared as described above may be subsequently engineered to obtain nucleic acids which include desired portions of the extended cDNA using conventional
10 techniques such as subcloning, PCR, or *in vitro* oligonucleotide synthesis. For example, nucleic acids which include only the full coding sequences (*i.e.* the sequences encoding the signal peptide and the mature protein remaining after the signal peptide is cleaved off) may be obtained using techniques known to those skilled in the art. Alternatively, conventional techniques may be applied to obtain nucleic acids which contain only the coding sequences
15 for the mature protein remaining after the signal peptide is cleaved off or nucleic acids which contain only the coding sequences for the signal peptides.

Similarly, nucleic acids containing any other desired portion of the coding sequences for the secreted protein may be obtained. For example, the nucleic acid may contain at least 10 consecutive bases of an extended cDNA such as one of the extended cDNAs described
20 below. In another embodiment, the nucleic acid may contain at least 15 consecutive bases of an extended cDNA such as one of the extended cDNAs described below. Alternatively, the nucleic acid may contain at least 20 consecutive bases of an extended cDNA such as one of the extended cDNAs described below. In another embodiment, the nucleic acid may contain at least 25 consecutive bases of an extended cDNA such as one of the extended cDNAs
25 described below. In yet another embodiment, the nucleic acid may contain at least 40 consecutive bases of an extended cDNA such as one of the extended cDNAs described below.

Once an extended cDNA has been obtained, it can be sequenced to determine the amino acid sequence it encodes. Once the encoded amino acid sequence has been
30 determined, one can create and identify any of the many conceivable cDNAs that will encode that protein by simply using the degeneracy of the genetic code. For example, allelic variants

or other homologous nucleic acids can be identified as described below. Alternatively, nucleic acids encoding the desired amino acid sequence can be synthesized *in vitro*.

In a preferred embodiment, the coding sequence may be selected using the known codon or codon pair preferences for the host organism in which the cDNA is to be expressed.

5 The extended cDNAs derived from the 5' ESTS of the present invention were obtained as described in Example 28 below.

EXAMPLE 28

Characterization of cloned extended cDNAs obtained using 5' ESTs

10 The procedure described in Example 27 above was used to obtain the extended cDNAs derived from the 5' ESTs of the present invention in a variety of tissues. The following list provides a few examples of thus obtained extended cDNAs.

Using this approach, the full length cDNA of SEQ ID NO:17 (internal identification number 48-19-3-G1-FL1) was obtained. This cDNA falls into the "EST-ext" category
15 described above and encodes the signal peptide MKKVLLITAILAVAVG (SEQ ID NO: 18) having a von Heijne score of 8.2.

The full length cDNA of SEQ ID NO:19 (internal identification number 58-34-2-E7-FL2) was also obtained using this procedure. This cDNA falls into the "EST-ext" category described above and encodes the signal peptide MWWFQQGLSFLPSALVTWTS (SEQ ID
20 NO:20) having a von Heijne score of 5.5.

Another full length cDNA obtained using the procedure described above has the sequence of SEQ ID NO:21 (internal identification number 51-27-1-E8-FL1). This cDNA, falls into the "EST-ext" category described above and encodes the signal peptide MVLTTLPANSANSPVNMPTTGPNLSYASSALSPCLT (SEQ ID NO:22) having a
25 von Heijne score of 5.9.

The above procedure was also used to obtain a full length cDNA having the sequence of SEQ ID NO:23 (internal identification number 76-4-1-G5-FL1). This cDNA falls into the "EST-ext" category described above and encodes the signal peptide ILSTVTALTFAXA (SEQ ID NO:24) having a von Heijne score of 5.5.

30 The full length cDNA of SEQ ID NO:25 (internal identification number 51-3-3-B10-FL3) was also obtained using this procedure. This cDNA falls into the "new" category

described above and encodes a signal peptide LVLTLCTLPLAVA (SEQ ID NO:26) having a von Heijne score of 10.1.

The full length cDNA of SEQ ID NO:27 (internal identification number 58-35-2-F10-FL2) was also obtained using this procedure. This cDNA falls into the "new" category
5 described above and encodes a signal peptide LWLLFFLVTAIHA (SEQ ID NO:28) having a von Heijne score of 10.7.

Bacterial clones containing plasmids containing the full length cDNAs described above are presently stored in the inventor's laboratories under the internal identification numbers provided above. The inserts may be recovered from the stored materials by growing
10 an aliquot of the appropriate bacterial clone in the appropriate medium. The plasmid DNA can then be isolated using plasmid isolation procedures familiar to those skilled in the art such as alkaline lysis minipreps or large scale alkaline lysis plasmid isolation procedures. If desired the plasmid DNA may be further enriched by centrifugation on a cesium chloride gradient, size exclusion chromatography, or anion exchange chromatography. The plasmid DNA
15 obtained using these procedures may then be manipulated using standard cloning techniques familiar to those skilled in the art. Alternatively, a PCR can be done with primers designed at both ends of the cDNA insertion. The PCR product which corresponds to the cDNA can then be manipulated using standard cloning techniques familiar to those skilled in the art.

The polypeptides encoded by the extended cDNAs may be screened for the presence
20 of known structural or functional motifs or for the presence of signatures, small amino acid sequences which are well conserved amongst the members of a protein family. The conserved regions have been used to derive consensus patterns or matrices included in the PROSITE data bank, in particular in the file prosite.dat (Release 13.0 of November 1995, located at <http://expasy.hcuge.ch/sprot/prosite.html>. Prosite_convert and prosite_scan
25 programs (http://ulrec3.unil.ch/ftpserveur/prosite_scan) may be used to find signatures on the extended cDNAs.

For each pattern obtained with the prosite_convert program from the prosite.dat file, the accuracy of the detection on a new protein sequence may be assessed by evaluating the frequency of irrelevant hits on the population of human secreted proteins included in the data
30 bank SWISSPROT. The ratio between the number of hits on shuffled proteins (with a window size of 20 amino acids) and the number of hits on native (unshuffled) proteins may be

used as an index. Every pattern for which the ratio is greater than 20% (one hit on shuffled proteins for 5 hits on native proteins) may be skipped during the search with `prosite_scan`. The program used to shuffle protein sequences (`db_shuffled`) and the program used to determine the statistics for each pattern in the protein data banks (`prosite_statistics`) are
5 available on the ftp site http://ulrec3.unil.ch/ftpserveur/prosite_scan.

In addition to PCR based methods for obtaining extended cDNAs, traditional hybridization based methods may also be employed. These methods may also be used to obtain the genomic DNAs which encode the mRNAs from which the 5' ESTs were derived,
10 mRNAs corresponding to the extended cDNAs, or nucleic acids which are homologous to extended cDNAs or 5' ESTs. Example 29 below provides examples of such methods.

EXAMPLE 29

Methods for Obtaining cDNAs which include the Entire Coding Region and the Authentic

15 5'End of the Corresponding mRNA

A full length cDNA library can be made using the strategies described in Examples 13, 14, 15, and 16 above by replacing the random nonamer used in Example 14 with an oligo-dT primer. For instance, the oligonucleotide of SEQ ID NO:14 may be used.

Alternatively, a cDNA library or genomic DNA library may be obtained from a
20 commercial source or made using techniques familiar to those skilled in the art. Such cDNA or genomic DNA libraries may be used to isolate extended cDNAs obtained from 5' EST or nucleic acids homologous to extended cDNAs or 5' EST as follows. The cDNA library or genomic DNA library is hybridized to a detectable probe comprising at least 10 consecutive nucleotides from the 5' EST or extended cDNA using conventional techniques. Preferably,
25 the probe comprises at least 12, 15, or 17 consecutive nucleotides from the 5' EST or extended cDNA. More preferably, the probe comprises at least 20 to 30 consecutive nucleotides from the 5' EST or extended cDNA. In some embodiments, the probe comprises more than 30 nucleotides from the 5' EST or extended cDNA.

Techniques for identifying cDNA clones in a cDNA library which hybridize to a given
30 probe sequence are disclosed in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*

2d Ed., Cold Spring Harbor Laboratory Press, 1989, the disclosure of which is incorporated herein by reference. The same techniques may be used to isolate genomic DNAs.

Briefly, cDNA or genomic DNA clones which hybridize to the detectable probe are identified and isolated for further manipulation as follows. A probe comprising at least 10 consecutive nucleotides from the 5' EST or extended cDNA is labeled with a detectable label such as a radioisotope or a fluorescent molecule. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the 5' EST or extended cDNA. More preferably, the probe comprises 20 to 30 consecutive nucleotides from the 5' EST or extended cDNA. In some embodiments, the probe comprises more than 30 nucleotides from the 5' EST or extended cDNA.

Techniques for labeling the probe are well known and include phosphorylation with polynucleotide kinase, nick translation, *in vitro* transcription, and non radioactive techniques. The cDNAs or genomic DNAs in the library are transferred to a nitrocellulose or nylon filter and denatured. After blocking of non specific sites, the filter is incubated with the labeled probe for an amount of time sufficient to allow binding of the probe to cDNAs or genomic DNAs containing a sequence capable of hybridizing thereto.

By varying the stringency of the hybridization conditions used to identify extended cDNAs or genomic DNAs which hybridize to the detectable probe, extended cDNAs having different levels of homology to the probe can be identified and isolated as described below.

1. Identification of Extended cDNA or Genomic cDNA Sequences Having a High Degree of Homology to the Labeled Probe

To identify extended cDNAs or genomic DNAs having a high degree of homology to the probe sequence, the melting temperature of the probe may be calculated using the following formulas:

For probes between 14 and 70 nucleotides in length the melting temperature (T_m) is calculated using the formula: $T_m = 81.5 + 16.6(\log [Na^+]) + 0.41(\text{fraction G+C}) - (600/N)$ where N is the length of the probe.

If the hybridization is carried out in a solution containing formamide, the melting temperature may be calculated using the equation $T_m = 81.5 + 16.6(\log [Na^+]) + 0.41(\text{fraction G+C}) - (0.63\% \text{ formamide}) - (600/N)$ where N is the length of the probe.

Prehybridization may be carried out in 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 µg denatured fragmented salmon sperm DNA or 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 100 µg denatured fragmented salmon sperm DNA, 50% formamide. The formulas for SSC and Denhardt's solutions are listed in Sambrook *et al.*, *supra*.

5 Hybridization is conducted by adding the detectable probe to the prehybridization solutions listed above. Where the probe comprises double stranded DNA, it is denatured before addition to the hybridization solution. The filter is contacted with the hybridization solution for a sufficient period of time to allow the probe to hybridize to extended cDNAs or genomic DNAs containing sequences complementary thereto or homologous thereto. For
10 probes over 200 nucleotides in length, the hybridization may be carried out at 15-25°C below the T_m. For shorter probes, such as oligonucleotide probes, the hybridization may be conducted at 15-25°C below the T_m. Preferably, for hybridizations in 6X SSC, the hybridization is conducted at approximately 68°C. Preferably, for hybridizations in 50% formamide containing solutions, the hybridization is conducted at approximately 42°C.

15 All of the foregoing hybridizations would be considered to be under "stringent" conditions.

Following hybridization, the filter is washed in 2X SSC, 0.1% SDS at room temperature for 15 minutes. The filter is then washed with 0.1X SSC, 0.5% SDS at room temperature for 30 minutes to 1 hour. Thereafter, the solution is washed at the hybridization
20 temperature in 0.1X SSC, 0.5% SDS. A final wash is conducted in 0.1X SSC at room temperature.

Extended cDNAs, nucleic acids homologous to extended cDNAs or 5' ESTs, or genomic DNAs which have hybridized to the probe are identified by autoradiography or other conventional techniques.

25 2. Obtention of Extended cDNA or Genomic cDNA Sequences Having Lower Degrees of Homology to the Labeled Probe

The above procedure may be modified to identify extended cDNAs, nucleic acids homologous to extended cDNAs, or genomic DNAs having decreasing levels of homology to the probe sequence. For example, to obtain extended cDNAs, nucleic acids homologous to
30 extended cDNAs, or genomic DNAs of decreasing homology to the detectable probe, less stringent conditions may be used. For example, the hybridization temperature may be

decreased in increments of 5°C from 68°C to 42°C in a hybridization buffer having a sodium concentration of approximately 1M. Following hybridization, the filter may be washed with 2X SSC, 0.5% SDS at the temperature of hybridization. These conditions are considered to be "moderate" conditions above 50°C and "low" conditions below 50°C.

- 5 Alternatively, the hybridization may be carried out in buffers, such as 6X SSC, containing formamide at a temperature of 42°C. In this case, the concentration of formamide in the hybridization buffer may be reduced in 5% increments from 50% to 0% to identify clones having decreasing levels of homology to the probe. Following hybridization, the filter may be washed with 6X SSC, 0.5% SDS at 50°C. These conditions are considered to be
- 10 "moderate" conditions above 25% formamide and "low" conditions below 25% formamide.

Extended cDNAs, nucleic acids homologous to extended cDNAs, or genomic DNAs which have hybridized to the probe are identified by autoradiography.

3. Determination of the Degree of Homology Between the Obtained Extended cDNAs and the Labeled Probe

- 15 If it is desired to obtain nucleic acids homologous to extended cDNAs, such as allelic variants thereof or nucleic acids encoding proteins related to the proteins encoded by the extended cDNAs, the level of homology between the hybridized nucleic acid and the extended cDNA or 5' EST used as the probe may be further determined using BLAST2N; parameters may be adapted depending on the sequence length and degree of homology
- 20 studied. To determine the level of homology between the hybridized nucleic acid and the extended cDNA or 5' EST from which the probe was derived, the nucleotide sequences of the hybridized nucleic acid and the extended cDNA or 5' EST from which the probe was derived are compared. For example, using the above methods, nucleic acids having at least 95% nucleic acid homology to the extended cDNA or 5' EST from which the probe was derived
- 25 may be obtained and identified. Similarly, by using progressively less stringent hybridization conditions one can obtain and identify nucleic acids having at least 90%, at least 85%, at least 80% or at least 75% homology to the extended cDNA or 5' EST from which the probe was derived.

- 30 To determine whether a clone encodes a protein having a given amount of homology to the protein encoded by the extended cDNA or 5' EST, the amino acid sequence encoded by the extended cDNA or 5' EST is compared to the amino acid sequence encoded by the

hybridizing nucleic acid. Homology is determined to exist when an amino acid sequence in the extended cDNA or 5' EST is closely related to an amino acid sequence in the hybridizing nucleic acid. A sequence is closely related when it is identical to that of the extended cDNA or 5' EST or when it contains one or more amino acid substitutions therein in which amino acids having similar characteristics have been substituted for one another. Using the above methods and algorithms such as FASTA with parameters depending on the sequence length and degree of homology studied, one can obtain nucleic acids encoding proteins having at least 95%, at least 90%, at least 85%, at least 80% or at least 75% homology to the proteins encoded by the extended cDNA or 5'EST from which the probe was derived.

In addition to the above described methods, other protocols are available to obtain extended cDNAs using 5' ESTs as outlined in the following paragraphs.

Extended cDNAs may be prepared by obtaining mRNA from the tissue, cell, or organism of interest using mRNA preparation procedures utilizing polyA selection procedures or other techniques known to those skilled in the art. A first primer capable of hybridizing to the polyA tail of the mRNA is hybridized to the mRNA and a reverse transcription reaction is performed to generate a first cDNA strand.

The first cDNA strand is hybridized to a second primer containing at least 10 consecutive nucleotides of the sequences of SEQ ID NOs 38-291. Preferably, the primer comprises at least 12, 15, or 17 consecutive nucleotides from the sequences of SEQ ID NOs 38-291. More preferably, the primer comprises 20 to 30 consecutive nucleotides from the sequences of SEQ ID NOs 38-291. In some embodiments, the primer comprises more than 30 nucleotides from the sequences of SEQ ID NOs 38-291. If it is desired to obtain extended cDNAs containing the full protein coding sequence, including the authentic translation initiation site, the second primer used contains sequences located upstream of the translation initiation site. The second primer is extended to generate a second cDNA strand complementary to the first cDNA strand. Alternatively, RT-PCR may be performed as described above using primers from both ends of the cDNA to be obtained.

Extended cDNAs containing 5' fragments of the mRNA may be prepared by hybridizing an mRNA comprising the sequence of the 5'EST for which an extended cDNA is desired with a primer comprising at least 10 consecutive nucleotides of the sequences

complementary to the 5'EST and reverse transcribing the hybridized primer to make a first cDNA strand from the mRNAs. Preferably, the primer comprises at least 12, 15, or 17 consecutive nucleotides from the 5'EST. More preferably, the primer comprises 20 to 30 consecutive nucleotides from the 5'EST.

5 Thereafter, a second cDNA strand complementary to the first cDNA strand is synthesized. The second cDNA strand may be made by hybridizing a primer complementary to sequences in the first cDNA strand to the first cDNA strand and extending the primer to generate the second cDNA strand.

 The double stranded extended cDNAs made using the methods described above are
10 isolated and cloned. The extended cDNAs may be cloned into vectors such as plasmids or viral vectors capable of replicating in an appropriate host cell. For example, the host cell may be a bacterial, mammalian, avian, or insect cell.

 Techniques for isolating mRNA, reverse transcribing a primer hybridized to mRNA to generate a first cDNA strand, extending a primer to make a second cDNA strand
15 complementary to the first cDNA strand, isolating the double stranded cDNA and cloning the double stranded cDNA are well known to those skilled in the art and are described in *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc. 1997 and Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, 1989, the entire disclosures of which are incorporated herein by reference.

20 Alternatively, procedures such as the one described in Example 29 may be used for obtaining full length cDNAs or extended cDNAs. In this approach, full length or extended cDNAs are prepared from mRNA and cloned into double stranded phagemids as follows. The cDNA library in the double stranded phagemids is then rendered single stranded by treatment with an endonuclease, such as the Gene II product of the phage F1, and an
25 exonuclease (Chang *et al.*, *Gene* 127:95-8, 1993). A biotinylated oligonucleotide comprising the sequence of a 5' EST, or a fragment containing at least 10 nucleotides thereof, is hybridized to the single stranded phagemids. Preferably, the fragment comprises at least 12, 15, or 17 consecutive nucleotides from the 5' EST. More preferably, the fragment comprises 20-30 consecutive nucleotides from the 5' EST. In some procedures, the fragment may
30 comprise more than 30 consecutive nucleotides from the 5' EST.

Hybrids between the biotinylated oligonucleotide and phagemids having inserts containing the 5' EST sequence are isolated by incubating the hybrids with streptavidin coated paramagnetic beads and retrieving the beads with a magnet (Fry *et al.*, *Biotechniques*, 13: 124-131, 1992). Thereafter, the resulting phagemids containing the 5' EST sequence are released from the beads and converted into double stranded DNA using a primer specific for the 5' EST sequence. Alternatively, protocols such as the Gene Trapper kit (Gibco BRL) may be used. The resulting double stranded DNA is transformed into bacteria. Extended cDNAs containing the 5' EST sequence are identified by colony PCR or colony hybridization.

Using any of the above described methods in section III, a plurality of extended cDNAs containing full length protein coding sequences or sequences encoding only the mature protein remaining after the signal peptide is cleaved off may be provided as cDNA libraries for subsequent evaluation of the encoded proteins or use in diagnostic assays as described below.

15

IV. Expression of Proteins Encoded by Extended cDNAs Isolated Using 5' ESTs

Extended cDNAs containing the full protein coding sequences of their corresponding mRNAs or portions thereof, such as cDNAs encoding the mature protein, may be used to express the encoded secreted proteins or portions thereof as described in Example 30 below. If desired, the extended cDNAs may contain the sequences encoding the signal peptide to facilitate secretion of the expressed protein. It will be appreciated that a plurality of extended cDNAs containing the full protein coding sequences or portions thereof may be simultaneously cloned into expression vectors to create an expression library for analysis of the encoded proteins as described below.

25

EXAMPLE 30

Expression of the Proteins Encoded by the Genes Corresponding to 5'ESTS or Portions Thereof

To express the proteins encoded by the genes corresponding to 5' ESTs (or portions thereof), full length cDNAs containing the entire protein coding region or extended cDNAs containing sequences adjacent to the 5' ESTs (or portions thereof) are obtained as described

30

in Examples 27-29 and cloned into a suitable expression vector. If desired, the nucleic acids may contain the sequences encoding the signal peptide to facilitate secretion of the expressed protein. The nucleic acids inserted into the expression vectors may also contain sequences upstream of the sequences encoding the signal peptide, such as sequences which regulate expression levels or sequences which confer tissue specific expression.

The nucleic acid encoding the protein or polypeptide to be expressed is operably linked to a promoter in an expression vector using conventional cloning technology. The expression vector may be any of the mammalian, yeast, insect or bacterial expression systems known in the art. Commercially available vectors and expression systems are available from a variety of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism in which the expression vector is introduced, as explained by Hatfield, *et al.*, U.S. Patent No. 5,082,767, incorporated herein by this reference.

The cDNA cloned into the expression vector may encode the entire protein (*i.e.* the signal peptide and the mature protein), the mature protein (*i.e.* the protein created by cleaving the signal peptide off), only the signal peptide or any other portion thereof.

The following is provided as one exemplary method to express the proteins encoded by the extended cDNAs corresponding to the 5' ESTs or the nucleic acids described above. First, the methionine initiation codon for the gene and the polyA signal of the gene are identified. If the nucleic acid encoding the polypeptide to be expressed lacks a methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the extended cDNA lacks a polyA signal, this sequence can be added to the construct by, for example, splicing out the polyA signal from pSG5 (Stratagene) using BglII and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the *gag* gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct allow efficient stable transfection. The vector includes the Herpes Simplex thymidine kinase promoter and the selectable neomycin gene. The extended cDNA or portion thereof encoding the polypeptide to be expressed is obtained

by PCR from the bacterial vector using oligonucleotide primers complementary to the extended cDNA or portion thereof and containing restriction endonuclease sequences for Pst I incorporated into the 5' primer and BglII at the 5' end of the corresponding cDNA 3' primer, taking care to ensure that the extended cDNA is positioned with the poly A signal. The
5 purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1 containing a poly A signal and prepared for this ligation (blunt/BglII).

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product
10 specification. Positive transfectants are selected after growing the transfected cells in 600 µg/ml G418 (Sigma, St. Louis, Missouri). Preferably the expressed protein is released into the culture medium, thereby facilitating purification.

Alternatively, the extended cDNAs may be cloned into pED6dpc2 as described above. The resulting pED6dpc2 constructs may be transfected into a suitable host cell, such
15 as COS 1 cells. Methotrexate resistant cells are selected and expanded. Preferably, the protein expressed from the extended cDNA is released into the culture medium thereby facilitating purification.

Proteins in the culture medium are separated by gel electrophoresis. If desired, the proteins may be ammonium sulfate precipitated or separated based on size or charge prior to
20 electrophoresis.

As a control, the expression vector lacking a cDNA insert is introduced into host cells or organisms and the proteins in the medium are harvested. The secreted proteins present in the medium are detected using techniques familiar to those skilled in the art such as Coomassie blue or silver staining or using antibodies against the protein encoded by the
25 extended cDNA

Antibodies capable of specifically recognizing the protein of interest may be generated using synthetic 15-mer peptides having a sequence encoded by the appropriate 5' EST, extended cDNA, or portion thereof. The synthetic peptides are injected into mice to generate antibody to the polypeptide encoded by the 5' EST, extended cDNA, or portion thereof.

30 Secreted proteins from the host cells or organisms containing an expression vector which contains the extended cDNA derived from a 5' EST or a portion thereof are compared

to those from the control cells or organism. The presence of a band in the medium from the cells containing the expression vector which is absent in the medium from the control cells indicates that the extended cDNA encodes a secreted protein. Generally, the band corresponding to the protein encoded by the extended cDNA will have a mobility near that expected based on the number of amino acids in the open reading frame of the extended cDNA. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

Alternatively, if the protein expressed from the above expression vectors does not contain sequences directing its secretion, the proteins expressed from host cells containing an expression vector with an insert encoding a secreted protein or portion thereof can be compared to the proteins expressed in control host cells containing the expression vector without an insert. The presence of a band in samples from cells containing the expression vector with an insert which is absent in samples from cells containing the expression vector without an insert indicates that the desired protein or portion thereof is being expressed. Generally, the band will have the mobility expected for the secreted protein or portion thereof. However, the band may have a mobility different than that expected as a result of modifications such as glycosylation, ubiquitination, or enzymatic cleavage.

The protein encoded by the extended cDNA may be purified using standard immunochromatography techniques. In such procedures, a solution containing the secreted protein, such as the culture medium or a cell extract, is applied to a column having antibodies against the secreted protein attached to the chromatography matrix. The secreted protein is allowed to bind the immunochromatography column. Thereafter, the column is washed to remove non-specifically bound proteins. The specifically bound secreted protein is then released from the column and recovered using standard techniques.

If antibody production is not possible, the extended cDNA sequence or portion thereof may be incorporated into expression vectors designed for use in purification schemes employing chimeric polypeptides. In such strategies, the coding sequence of the extended cDNA or portion thereof is inserted in frame with the gene encoding the other half of the chimera. The other half of the chimera may be β -globin or a nickel binding polypeptide. A chromatography matrix having antibody to β -globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites may be engineered between the β -globin

gene or the nickel binding polypeptide and the extended cDNA or portion thereof. Thus, the two polypeptides of the chimera may be separated from one another by protease digestion.

One useful expression vector for generating β -globin chimerics is pSG5 (Stratagene), which encodes rabbit β -globin. Intron II of the rabbit β -globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques as described are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis *et al.*, (*Basic Methods in Molecular Biology*, Davis, DIBNER, and BATTEY, ed., Elsevier Press, NY, 1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from the construct using *in vitro* translation systems such as the *In vitro Express*TM Translation Kit (Stratagene).

Following expression and purification of the secreted proteins encoded by the 5' ESTs, extended cDNAs, or fragments thereof, the purified proteins may be tested for the ability to bind to the surface of various cell types as described in Example 31 below. It will be appreciated that a plurality of proteins expressed from these cDNAs may be included in a panel of proteins to be simultaneously evaluated for the activities specifically described below, as well as other biological roles for which assays for determining activity are available.

20

EXAMPLE 31

Analysis of Secreted Proteins to Determine Whether they Bind to the Cell Surface

The proteins encoded by the 5' ESTs, extended cDNAs, or fragments thereof are cloned into expression vectors such as those described in Example 30. The proteins are purified by size, charge, immunochromatography or other techniques familiar to those skilled in the art. Following purification, the proteins are labeled using techniques known to those skilled in the art. The labeled proteins are incubated with cells or cell lines derived from a variety of organs or tissues to allow the proteins to bind to any receptor present on the cell surface. Following the incubation, the cells are washed to remove non-specifically bound protein. The labeled proteins are detected by autoradiography. Alternatively, unlabeled proteins may be incubated with the cells and detected with antibodies having a detectable label, such as a fluorescent molecule, attached thereto.

Specificity of cell surface binding may be analyzed by conducting a competition analysis in which various amounts of unlabeled protein are incubated along with the labeled protein. The amount of labeled protein bound to the cell surface decreases as the amount of competitive unlabeled protein increases. As a control, various amounts of an unlabeled protein unrelated to the labeled protein is included in some binding reactions. The amount of labeled protein bound to the cell surface does not decrease in binding reactions containing increasing amounts of unrelated unlabeled protein, indicating that the protein encoded by the cDNA binds specifically to the cell surface.

As discussed above, secreted proteins have been shown to have a number of important physiological effects and, consequently, represent a valuable therapeutic resource. The secreted proteins encoded by the extended cDNAs or portions thereof made according to Examples 27-29 may be evaluated to determine their physiological activities as described below.

EXAMPLE 32

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Cytokine, Cell Proliferation or Cell Differentiation Activity

As discussed above, secreted proteins may act as cytokines or may affect cellular proliferation or differentiation. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor dependent cell proliferation assays, and hence the assays serve as a convenient confirmation of cytokine activity. The activity of a protein encoded by the extended cDNAs is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M⁺ (preB M⁺), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7c and CMK. The proteins encoded by the above extended cDNAs or portions thereof may be evaluated for their ability to regulate T cell or thymocyte proliferation in assays such as those described above or in the following references, which are incorporated herein by reference: *Current Protocols in Immunology*, Ed. by Coligan *et al.*, Greene Publishing Associates and Wiley-Interscience; Takai *et al. J. Immunol.* 137:3494-3500, 1986., Bertagnolli *et al., J. Immunol.* 145:1706-1712, 1990., Bertagnolli *et al., Cell*

Immunol. **133**:327-341, 1991; Bertagnolli, *et al.*, *J. Immunol.* **149**:3778-3783, 1992; Bowman *et al.*, *J. Immunol.* **152**:1756-1761, 1994.

In addition, numerous assays for cytokine production and/or the proliferation of spleen cells, lymph node cells and thymocytes are known. These include the techniques disclosed in *Current Protocols in Immunology, supra* 1:3.12.1-3.12.14; and Schreiber In *Current Protocols in Immunology, supra* 1 : 6.8.1-6.8.8.

The proteins encoded by the cDNAs may also be assayed for the ability to regulate the proliferation and differentiation of hematopoietic or lymphopoietic cells. Many assays for such activity are familiar to those skilled in the art, including the assays in the following references, which are incorporated herein by reference: Bottomly *et al.*, In *Current Protocols in Immunology, supra* 1 : 6.3.1-6.3.12.; deVries *et al.*, *J. Exp. Med.* **173**:1205-1211, 1991; Moreau *et al.*, *Nature* **36**:690-692, 1988; Greenberger *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **80**:2931-2938, 1983; Nordan, R., In *Current Protocols in Immunology, supra* 1 : 6.6.1-6.6.5; Smith *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**:1857-1861, 1986; Bennett *et al.*, in *Current Protocols in Immunology supra* 1 : 6.15.1; Ciarletta *et al.*, In *Current Protocols in Immunology, supra* 1 : 6.13.1.

The proteins encoded by the cDNAs may also be assayed for their ability to regulate T-cell responses to antigens. Many assays for such activity are familiar to those skilled in the art, including the assays described in the following references, which are incorporated herein by reference: Chapter 3 (*In Vitro* Assays for Mouse Lymphocyte Function), Chapter 6 (Cytokines and Their Cellular Receptors) and Chapter 7, (Immunologic Studies in Humans) in *Current Protocols in Immunology supra*; Weinberger *et al.*, *Proc. Natl. Acad. Sci. USA* **77**:6091-6095, 1980; Weinberger *et al.*, *Eur. J. Immun.* **11**:405-411, 1981; Takai *et al.*, *J. Immunol.* **137**:3494-3500, 1986; Takai *et al.*, *J. Immunol.* **140**:508-512, 1988.

Those proteins which exhibit cytokine, cell proliferation, or cell differentiation activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which induction of cell proliferation or differentiation is beneficial. Alternatively, as described in more detail below, genes encoding these proteins or nucleic acids regulating the expression of these proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 33

Assaying the Proteins Expressed from Extended cDNAs or PortionsThereof for Activity as Immune System Regulators

The proteins encoded by the cDNAs may also be evaluated for their effects as
5 immune regulators. For example, the proteins may be evaluated for their activity to influence
thymocyte or splenocyte cytotoxicity. Numerous assays for such activity are familiar to those
skilled in the art including the assays described in the following references, which are
incorporated herein by reference: Chapter 3 (*In Vitro* Assays for Mouse Lymphocyte
Function 3.1-3.19) and Chapter 7 (Immunologic studies in Humans) in *Current Protocols in*
10 *Immunology*, Coligan *et al.*, Eds, Greene Publishing Associates and Wiley-Interscience;
Herrmann *et al.*, *Proc. Natl. Acad. Sci. USA* 78:2488-2492, 1981; Herrmann *et al.*, *J.*
Immunol. 128:1968-1974, 1982; Handa *et al.*, *J. Immunol.* 135:1564-1572, 1985; Takai *et*
al., *J. Immunol.* 137:3494-3500, 1986; Takai *et al.*, *J. Immunol.* 140:508-512, 1988;
Bowman *et al.*, *J. Virology* 61:1992-1998; Bertagnolli *et al.*, *Cell. Immunol.* 133:327-341,
15 1991; Brown *et al.*, *J. Immunol.* 153:3079-3092, 1994.

The proteins encoded by the cDNAs may also be evaluated for their effects on T-cell
dependent immunoglobulin responses and isotype switching. Numerous assays for such
activity are familiar to those skilled in the art, including the assays disclosed in the following
references, which are incorporated herein by reference: Maliszewski, *J. Immunol.* 144:3028-
20 3033, 1990; Mond *et al.* in *Current Protocols in Immunology*, 1 : 3.8.1-3.8.16, *supra*.

The proteins encoded by the cDNAs may also be evaluated for their effect on immune
effector cells, including their effect on Th1 cells and cytotoxic lymphocytes. Numerous assays
for such activity are familiar to those skilled in the art, including the assays disclosed in the
following references, which are incorporated herein by reference: Chapter 3 (*In Vitro* Assays
25 for Mouse Lymphocyte Function 3.1-3.19) and Chapter 7 (Immunologic Studies in Humans)
in *Current Protocols in Immunology*, *supra*; Takai *et al.*, *J. Immunol.* 137:3494-3500, 1986;
Takai *et al.*, *J. Immunol.* 140:508-512, 1988; Bertagnolli *et al.*, *J. Immunol.* 149:3778-3783,
1992.

The proteins encoded by the cDNAs may also be evaluated for their effect on
30 dendritic cell mediated activation of naive T-cells. Numerous assays for such activity are
familiar to those skilled in the art, including the assays disclosed in the following references,

which are incorporated herein by reference: Guery *et al.*, *J. Immunol.* 134:536-544, 1995; Inaba *et al.*, *J. Exp. Med.* 173:549-559, 1991; Macatonia *et al.*, *J. Immunol.* 154:5071-5079, 1995; Porgador *et al.*, *J. Exp. Med.* 182:255-260, 1995; Nair *et al.*, *J. Virol.* 67:4062-4069, 1993; Huang *et al.*, *Science* 264:961-965, 1994; Macatonia *et al.*, *J. Exp. Med.* 169:1255-1264, 1989; Bhardwaj *et al.*, *Journal of Clinical Investigation* 94:797-807, 1994; and Inaba *et al.*, *J. Exp. Med.* 172:631-640, 1990.

The proteins encoded by the cDNAs may also be evaluated for their influence on the lifetime of lymphocytes. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Darzynkiewicz *et al.*, *Cytometry* 13:795-808, 1992; Gorczyca *et al.*, *Leukemia* 7:659-670, 1993; Gorczyca *et al.*, *Cancer Res.* 53:1945-1951, 1993; Itoh *et al.*, *Cell* 66:233-243, 1991; Zacharchuk, *J. Immunol.* 145:4037-4045, 1990; Zamai *et al.*, *Cytometry* 14:891-897, 1993; Gorczyca *et al.*, *Int. J. Oncol.* 1:639-648, 1992.

The proteins encoded by the cDNAs may also be evaluated for their influence on early steps of T-cell commitment and development. Numerous assays for such activity are familiar to those skilled in the art, including without limitation the assays disclosed in the following references, which are incorporated herein by references: Antica *et al.*, *Blood* 84:111-117, 1994; Fine *et al.*, *Cell. Immunol.* 155:111-122, 1994; Galy *et al.*, *Blood* 85:2770-2778, 1995; Toki *et al.*, *Proc. Nat. Acad. Sci. USA* 88:7548-7551, 1991.

Those proteins which exhibit activity as immune system regulators activity may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of immune activity is beneficial. For example, the protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases caused by viral, bacterial, fungal or other infection may be treatable using a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention, including infections by HIV, hepatitis viruses, herpesviruses, mycobacteria, *Leishmania* spp., *Plasmodium* and various fungal infections such as candidiasis. Of course, in this regard, a protein encoded by

extended cDNAs derived from the 5' ESTs of the present invention may also be useful where a boost to the immune system generally may be desirable, *i.e.*, in the treatment of cancer.

Alternatively, proteins encoded by extended cDNAs derived from the 5' ESTs of the present invention may be used in treatment of autoimmune disorders including, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitus, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also to be useful in the treatment of allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention.

Using the proteins of the invention it may also be possible to regulate immune responses either up or down.

Down regulation may involve inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T-cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active non-antigen-specific process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after the end of exposure to the tolerizing agent. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions, such as, for example, B7 costimulation), *e.g.*, preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through

its recognition as foreign by T cells, followed by an immune reaction that destroys the transplant. The administration of a molecule which inhibits or blocks interaction of a B lymphocyte antigen with its natural ligand(s) on immune cells (such as a soluble, monomeric form of a peptide having B7-2 activity alone or in conjunction with a monomeric form of a peptide having an activity of another B lymphocyte antigen (e.g., B7-1, B7-3) or blocking antibody), prior to transplantation, can lead to the binding of the molecule to the natural ligand(s) on the immune cells without transmitting the corresponding costimulatory signal. Blocking B lymphocyte antigen function in this matter prevents cytokine synthesis by immune cells, such as T cells, and thus acts as an immunosuppressant. Moreover, the lack of costimulation may also be sufficient to anergize the T cells, thereby inducing tolerance in a subject. Induction of long-term tolerance by B lymphocyte antigen-blocking reagents may avoid the necessity of repeated administration of these blocking reagents. To achieve sufficient immunosuppression or tolerance in a subject, it may also be necessary to block the function of a combination of B lymphocyte antigens.

The efficacy of particular blocking reagents in preventing organ transplant rejection or GVHD can be assessed using animal models that are predictive of efficacy in humans. Examples of appropriate systems which can be used include allogeneic cardiac grafts in rats and xenogeneic pancreatic islet cell grafts in mice, both of which have been used to examine the immunosuppressive effects of CTLA4Ig fusion proteins *in vivo* as described in Lenschow *et al.*, *Science* 257:789-792, 1992 and Turka *et al.*, *Proc. Natl. Acad. Sci USA*, 89:11102-11105, 1992. In addition, murine models of GVHD (see Paul ed., *Fundamental Immunology*, Raven Press, New York, 1989, pp. 846-847) can be used to determine the effect of blocking B lymphocyte antigen function *in vivo* on the development of that disease.

Blocking antigen function may also be therapeutically useful for treating autoimmune diseases. Many autoimmune disorders are the result of inappropriate activation of T cells that are reactive against self tissue and which promote the production of cytokines and autoantibodies involved in the pathology of the diseases. Preventing the activation of autoreactive T cells may reduce or eliminate disease symptoms. Administration of reagents which block costimulation of T cells by disrupting receptor/ligand interactions of B lymphocyte antigens can be used to inhibit T cell activation and prevent production of autoantibodies or T cell-derived cytokines which potentially involved in the disease process.

Additionally, blocking reagents may induce antigen-specific tolerance of autoreactive T cells which could lead to long-term relief from the disease. The efficacy of blocking reagents in preventing or alleviating autoimmune disorders can be determined using a number of well-characterized animal models of human autoimmune diseases. Examples include murine
5 experimental autoimmune encephalitis, systemic lupus erythematosus in MRL/pr/pr mice or NZB hybrid mice, murine autoimmune collagen arthritis, diabetes mellitus in OD mice and BB rats, and murine experimental myasthenia gravis (see Paul ed., *supra*, pp. 840-856).

Upregulation of an antigen function (preferably a B lymphocyte antigen function), as a means of up regulating immune responses, may also be useful in therapy. Upregulation of
10 immune responses may involve either enhancing an existing immune response or eliciting an initial immune response as shown by the following examples. For instance, enhancing an immune response through stimulating B lymphocyte antigen function may be useful in cases of viral infection. In addition, systemic viral diseases such as influenza, the common cold, and encephalitis might be alleviated by the administration of stimulatory form of B lymphocyte
15 antigens systemically.

Alternatively, antiviral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells *in vitro* with viral antigen-pulsed APCs either expressing a peptide encoded by extended cDNAs derived from the 5' ESTs of the present invention or together with a stimulatory form of a soluble peptide encoded by
20 extended cDNAs derived from the 5' ESTs of the present invention and reintroducing the *in vitro* primed T cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to T cells *in vivo*, thereby activating the T cells.

In another application, upregulation or enhancement of antigen function (preferably B lymphocyte antigen function) may be useful in the induction of tumor immunity. Tumor cells
25 (e.g., sarcoma, melanoma, lymphoma, leukemia, neuroblastoma, carcinoma) transfected with a nucleic acid encoding at least one peptide encoded by extended cDNAs derived from the 5' ESTs of the present invention can be administered to a subject to overcome tumor-specific tolerance in the subject. If desired, the tumor cell can be transfected to express a combination of peptides. For example, tumor cells obtained from a patient can be transfected *ex vivo* with
30 an expression vector directing the expression of a peptide having B7-2-like activity alone, or in conjunction with a peptide having B7-1-like activity and/or B7-3-like activity. The

transfected tumor cells are returned to the patient to result in expression of the peptides on the surface of the transfected cell. Alternatively, gene therapy techniques can be used to target a tumor cell for transfection *in vivo*.

The presence of the peptide encoded by extended cDNAs derived from the 5' ESTs of the present invention having the activity of a B lymphocyte antigen(s) on the surface of the tumor cell provides the necessary costimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack or which fail to reexpress sufficient amounts of MHC class I or MHC class II molecules can be transfected with nucleic acids encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I α chain and β_2 microglobulin or an MHC class II α chain and an MHC class II β chain to thereby express MHC class I or MHC class II proteins on the cell surface, respectively. Expression of the appropriate MHC class I or class II molecules in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject. Alternatively, as described in more detail below, genes encoding these immune system regulator proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

25

EXAMPLE 34

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Hematopoiesis Regulating Activity

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their hematopoiesis regulating activity. For example, the effect of the proteins on embryonic stem cell differentiation may be evaluated. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following

30

references, which are incorporated herein by reference: Johansson *et al.* *Cell. Biol.* 15:141-151, 1995; Keller *et al.*, *Mol. Cell. Biol.* 13:473-486, 1993; McClanahan *et al.*, *Blood* 81:2903-2915, 1993.

The proteins encoded by the extended cDNAs or portions thereof may also be
5 evaluated for their influence on the lifetime of stem cells and stem cell differentiation. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Freshney, Methylcellulose Colony Forming Assays, in *Culture of Hematopoietic Cells*, Freshney, *et al.* Eds. pp. 265-268, Wiley-Liss, Inc., New York, NY. 1994; Hirayama *et al.*, *Proc. Natl. Acad.*
10 *Sci. USA* 89:5907-5911, 1992; McNiece and Briddell, in *Culture of Hematopoietic Cells*, *supra*; Neben *et al.*, *Exp. Hematol.* 22:353-359, 1994; Ploemacher and Cobblestone In *Culture of Hematopoietic Cells*, *supra*1-21, Spooncer *et al.*, in *Culture of Hematopoietic Cells*, *supra*163-179 and Sutherland in *Culture of Hematopoietic Cells*, *supra*. 139-162.

Those proteins which exhibit hematopoiesis regulatory activity may then be
15 formulated as pharmaceuticals and used to treat clinical conditions in which regulation of hematopoiesis is beneficial, such as in the treatment of myeloid or lymphoid cell deficiencies. Involvement in regulating hematopoiesis is indicated even by marginal biological activity in support of colony forming cells or of factor-dependent cell lines. For example, proteins supporting the growth and proliferation of erythroid progenitor cells alone or in combination
20 with other cytokines, indicates utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells. Proteins supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (*i.e.*, traditional CSF activity) may be useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-
25 suppression. Proteins supporting the growth and proliferation of megakaryocytes and consequently of platelets allows prevention or treatment of various platelet disorders such as thrombocytopenia, and generally may be used in place of or complementary to platelet transfusions. Proteins supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells may
30 therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof
for Regulation of Tissue Growth

Assays for wound healing activity include, without limitation, those described in: Winter, *Epidermal Wound Healing*, pps. 71-112, Maibach and Rovee, eds., Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, *J. Invest. Dermatol.* 71:382-84, 1978, which are incorporated herein by reference.

A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention, which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Such a preparation employing a protein of the invention may have prophylactic use in closed as well as open fracture reduction and also in the

improved fixation of artificial joints. *De novo* bone synthesis induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

5 A protein of this invention may also be used in the treatment of periodontal disease, and in other tooth repair processes. Such agents may provide an environment to attract bone-forming cells, stimulate growth of bone-forming cells or induce differentiation of bone-forming cell progenitors. A protein of the invention may also be useful in the treatment of osteoporosis or osteoarthritis, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast
10 activity, etc.) mediated by inflammatory processes.

Another category of tissue regeneration activity that may be attributable to the protein encoded by extended cDNAs derived from the 5' ESTs of the present invention is tendon/ligament formation. A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention, which induces tendon/ligament-like tissue or other tissue
15 formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues,
20 and in repairing defects to tendon or ligament tissue. *De novo* tendon/ligament-like tissue formation induced by a composition encoded by extended cDNAs derived from the 5' ESTs of the present invention contributes to the repair of tendon or ligaments defects of congenital, traumatic or other origin and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions encoded by extended cDNAs derived from the 5'
25 ESTs of the present invention may provide an environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other
30 tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, *i.e.*, for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a protein may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a protein of the invention.

Proteins of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

It is expected that a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also exhibit activity for generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium) muscle (smooth, skeletal or cardiac) and vascular (including vascular endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring to allow normal tissue to generate. A protein of the invention may also exhibit angiogenic activity.

A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Alternatively, as described in more detail below, genes encoding tissue growth regulating activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

5

EXAMPLE 36

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Regulation of Reproductive Hormones

The proteins encoded by the extended cDNAs or portions thereof may also be
10 evaluated for their ability to regulate reproductive hormones, such as follicle stimulating hormone. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Vale *et al.*, *Endocrinol.* **91**:562-572, 1972; Ling *et al.*, *Nature* **321**:779-782, 1986; Vale *et al.*, *Nature* **321**:776-779, 1986; Mason *et al.*, *Nature* **318**:659-663, 1985; Forage *et al.*,
15 *Proc. Natl. Acad. Sci. USA* **83**:3091-3095, 1986, Chapter 6.12 in *Current Protocols in Immunology*, Coligan *et al.* Eds. Greene Publishing Associates and Wiley-Interscience; Taub *et al.*, *J. Clin. Invest.* **95**:1370-1376, 1995; Lind *et al.*, *APMIS* **103**:140-146, 1995; Muller *et al.*, *Eur. J. Immunol.* **25**:1744-1748; Gruber *et al.*, *J. Immunol.* **152**:5860-5867, 1994; Johnston *et al.*, *J Immunol.* **153**:1762-1768, 1994.

20 Those proteins which exhibit activity as reproductive hormones or regulators of cell movement may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of reproductive hormones are beneficial. For example, a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may also exhibit activin- or inhibin-related activities. Inhibins are characterized by their ability to inhibit the release of
25 follicle stimulating hormone (FSH), while activins are characterized by their ability to stimulate the release of FSH. Thus, a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention, alone or in heterodimers with a member of the inhibin α family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of
30 sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the protein of the invention, as a homodimer or as a heterodimer with other protein subunits of

the inhibin-B group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, United States Patent 4,798,885, the disclosure of which is incorporated herein by reference. A protein of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as cows, sheep and pigs.

Alternatively, as described in more detail below, genes encoding reproductive hormone regulating activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 37

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Chemotactic/Chemokinetic Activity

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for chemotactic/chemokinetic activity. For example, a protein encoded by extended cDNAs derived from the 5' ESTs of the present invention may have chemotactic or chemokinetic activity (e.g., act as a chemokine) for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. Chemotactic and chemokinetic proteins can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic proteins provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population. Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: *Current Protocols in Immunology*, Ed by Coligan, Kruisbeek, Margulies, Shevach and Strober, Pub. Greene Publishing Associates and Wiley-Interscience, Chapter 6.12: 6.12.1-6.12.28; Taub *et al.*, *J. Clin. Invest.* 95:1370-1376, 1995; Lind *et al.*, *APMIS* 103:140-146, 1995; Mueller *et al.*, *Eur. J. Immunol.* 25:1744-1748; Gruber *et al.*, *J. Immunol.* 152:5860-5867, 1994; Johnston *et al.* *J. Immunol.*, 153:1762-1768, 1994.

EXAMPLE 38

Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof for Regulation of Blood Clotting

The proteins encoded by the extended cDNAs or portions thereof may also be evaluated for their effects on blood clotting. Numerous assays for such activity are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Linet *et al.*, *J. Clin. Pharmacol.* 26:131-140, 1986; Burdick *et al.*, *Thrombosis Res.* 45:413-419, 1987; Humphrey *et al.*, *Fibrinolysis* 5:71-79, 1991; Schaub, *Prostaglandins* 35:467-474, 1988.

Those proteins which are involved in the regulation of blood clotting may then be formulated as pharmaceuticals and used to treat clinical conditions in which regulation of blood clotting is beneficial. For example, a protein of the invention may also exhibit hemostatic or thrombolytic activity. As a result, such a protein is expected to be useful in treatment of various coagulations disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A protein of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as infarction of cardiac and central nervous system

vessels (e.g., stroke)). Alternatively, as described in more detail below, genes encoding blood clotting activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

5

EXAMPLE 39

Assaying the Proteins Expressed from Extended cDNAs or

Portions Thereof for Involvement in Receptor/Ligand Interactions

The proteins encoded by the extended cDNAs or a portion thereof may also be
10 evaluated for their involvement in receptor/ligand interactions. Numerous assays for such involvement are familiar to those skilled in the art, including the assays disclosed in the following references, which are incorporated herein by reference: Chapter 7. 7.28.1-7.28.22 in *Current Protocols in Immunology*, Coligan *et al.* Eds. Greene Publishing Associates and Wiley-Interscience; Takai *et al.*, *Proc. Natl. Acad. Sci. USA* 84:6864-6868, 1987; Bierer *et al.*, *J. Exp. Med.* 168:1145-1156, 1988; Rosenstein *et al.*, *J. Exp. Med.* 169:149-160, 1989;
15 Stoltenborg *et al.*, *J. Immunol. Methods* 175:59-68, 1994; Stitt *et al.*, *Cell* 80:661-670, 1995; Gyuris *et al.*, *Cell* 75:791-803, 1993.

For example, the proteins encoded by extended cDNAs derived from the 5' ESTs of the present invention may also demonstrate activity as receptors, receptor ligands or inhibitors
20 or agonists of receptor/ligand interactions. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen
25 recognition and development of cellular and humoral immune responses). Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein encoded by extended cDNAs derived from the 5' ESTs of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions. Alternatively,
30 as described in more detail below, genes encoding proteins involved in receptor/ligand

interactions or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 40

5 Assaying the Proteins Expressed from Extended cDNAs or Portions Thereof
 for Anti-Inflammatory Activity

The proteins encoded by the extended cDNAs or a portion thereof may also be evaluated for anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or
10 promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Proteins exhibiting such activities can be used to treat inflammatory conditions including chronic or acute conditions, including
15 without limitation inflammation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine- or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Proteins of the invention may also be useful to
20 treat anaphylaxis and hypersensitivity to an antigenic substance or material. Alternatively, as described in more detail below, genes encoding anti-inflammatory activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

25 **EXAMPLE 41**

Assaying the Proteins Expressed from Extended cDNAs or
 Portions Thereof for Tumor Inhibition Activity

The proteins encoded by the extended cDNAs or a portion thereof may also be evaluated for tumor inhibition activity. In addition to the activities described above for
30 immunological treatment or prevention of tumors, a protein of the invention may exhibit other anti-tumor activities. A protein may inhibit tumor growth directly or indirectly (such as, for

example, via ADCC). A protein may exhibit its tumor inhibitory activity by acting on tumor tissue or tumor precursor tissue, by inhibiting formation of tissues necessary to support tumor growth (such as, for example, by inhibiting angiogenesis), by causing production of other factors, agents or cell types which inhibit tumor growth, or by suppressing, eliminating or
5 inhibiting factors, agents or cell types which promote tumor growth. Alternatively, as described in more detail below, genes tumor inhibition activity proteins or nucleic acids regulating the expression of such proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

10 A protein of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or
15 body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional factors or component(s); effecting behavioral
20 characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating
25 deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein. Alternatively, as described in more detail below, genes encoding proteins involved in
30 any of the above mentioned activities or nucleic acids regulating the expression of such

proteins may be introduced into appropriate host cells to increase or decrease the expression of the proteins as desired.

EXAMPLE 42

5 Identification of Proteins which Interact with
 Polypeptides Encoded by Extended cDNAs

Proteins which interact with the polypeptides encoded by cDNAs derived from the 5' ESTs or fragments thereof, such as receptor proteins, may be identified using two hybrid systems such as the Matchmaker Two Hybrid System 2 (Catalog No. K1604-1, Clontech).
10 As described in the manual accompanying the kit which is incorporated herein by reference, the the cDNAs derived from 5' ESTs, or fragments thereof, are inserted into an expression vector such that they are in frame with DNA encoding the DNA binding domain of the yeast transcriptional activator GAL4. cDNAs in a cDNA library which encode proteins which might interact with the polypeptides encoded by the extended cDNAs or portions thereof are
15 inserted into a second expression vector such that they are in frame with DNA encoding the activation domain of GAL4. The two expression plasmids are transformed into yeast and the yeast are plated on selection medium which selects for expression of selectable markers on each of the expression vectors as well as GAL4 dependent expression of the HIS3 gene. Transformants capable of growing on medium lacking histidine are screened for GAL4
20 dependent lacZ expression. Those cells which are positive in both the histidine selection and the lacZ assay contain plasmids encoding proteins which interact with the polypeptide encoded by the extended cDNAs or portions thereof.

Alternatively, the system described in Lustig *et al.*, *Methods in Enzymology* 283: 83-99, 1997, and in U.S. Patent No. 5,654,150, the disclosure of which is incorporated herein by
25 reference, may be used for identifying molecules which interact with the polypeptides encoded by extended cDNAs. In such systems, *in vitro* transcription reactions are performed on a pool of vectors containing extended cDNA inserts cloned downstream of a promoter which drives *in vitro* transcription. The resulting pools of mRNAs are introduced into *Xenopus laevis* oocytes. The oocytes are then assayed for a desired activity.

Alternatively, the pooled *in vitro* transcription products produced as described above may be translated *in vitro*. The pooled *in vitro* translation products can be assayed for a desired activity or for interaction with a known polypeptide.

Proteins or other molecules interacting with polypeptides encoded by extended
5 cDNAs can be found by a variety of additional techniques. In one method, affinity columns containing the polypeptide encoded by the extended cDNA or a portion thereof can be constructed. In some versions, of this method the affinity column contains chimeric proteins in which the protein encoded by the extended cDNA or a portion thereof is fused to glutathione S-transferase. A mixture of cellular proteins or pool of
10 expressed proteins as described above and is applied to the affinity column. Proteins interacting with the polypeptide attached to the column can then be isolated and analyzed on 2-D electrophoresis gel as described in Ramunsen *et al.*, *Electrophoresis* 18:588-598, 1997, the disclosure of which is incorporated herein by reference. Alternatively, the proteins retained on the affinity column can be purified by electrophoresis based methods
15 and sequenced. The same method can be used to isolate antibodies, to screen phage display products, or to screen phage display human antibodies.

Proteins interacting with polypeptides encoded by extended cDNAs or portions thereof can also be screened by using an Optical Biosensor as described in Edwards and Leatherbarrow, *Analytical Biochemistry* 246:1-6, 1997, the disclosure of which is
20 incorporated herein by reference. The main advantage of the method is that it allows the determination of the association rate between the protein and other interacting molecules. Thus, it is possible to specifically select interacting molecules with a high or low association rate. Typically a target molecule is linked to the sensor surface (through a carboxymethyl dextran matrix) and a sample of test molecules is placed in contact with
25 the target molecules. The binding of a test molecule to the target molecule causes a change in the refractive index and/ or thickness. This change is detected by the Biosensor provided it occurs in the evanescent field (which extend a few hundred nanometers from the sensor surface). In these screening assays, the target molecule can be one of the polypeptides encoded by extended cDNAs or a portion thereof and the test
30 sample can be a collection of proteins extracted from tissues or cells, a pool of expressed proteins, combinatorial peptide and/ or chemical libraries, or phage displayed peptides.

The tissues or cells from which the test proteins are extracted can originate from any species.

In other methods, a target protein is immobilized and the test population is a collection of unique polypeptides encoded by the extended cDNAs or portions thereof.

5 To study the interaction of the proteins encoded by the extended cDNAs or portions thereof with drugs, the microdialysis coupled to HPLC method described by Wang *et al.*, *Chromatographia* 44:205-208, 1997 or the affinity capillary electrophoresis method described by Busch *et al.*, *J. Chromatogr.* 777:311-328, 1997, the disclosures of which are incorporated herein by reference can be used.

10

It will be appreciated by those skilled in the art that the proteins expressed from the extended cDNAs or portions may be assayed for numerous activities in addition to those specifically enumerated above. For example, the expressed proteins may be evaluated for applications involving control and regulation of inflammation, tumor proliferation or
15 metastasis, infection, or other clinical conditions. In addition, the proteins expressed from the extended cDNAs or portions thereof may be useful as nutritional agents or cosmetic agents.

The proteins expressed from the cDNAs or portions thereof may be used to generate antibodies capable of specifically binding to the expressed protein or fragments thereof as described in Example 40 below. The antibodies may capable of binding a full length protein
20 encoded by a cDNA derived from a 5' EST, a mature protein (*i.e.* the protein generated by cleavage of the signal peptide) encoded by a cDNA derived from a 5' EST, or a signal peptide encoded by a cDNA derived from a 5' EST. Alternatively, the antibodies may be capable of binding fragments of at least 10 amino acids of the proteins encoded by the above cDNAs. In some embodiments, the antibodies may be capable of binding fragments of at
25 least 15 amino acids of the proteins encoded by the above cDNAs. In other embodiments, the antibodies may be capable of binding fragments of at least 25 amino acids of the proteins expressed from the extended cDNAs which comprise at least 25 amino acids of the proteins encoded by the above cDNAs. In further embodiments, the antibodies may be capable of binding fragments of at least 40 amino acids of the proteins encoded by the above cDNAs.

30

EXAMPLE 43

Production of an Antibody to a Human Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells as described in Example 30. The concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few $\mu\text{g/ml}$. Monoclonal or polyclonal antibody to the protein can then be prepared as follows:

1. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, and Milstein, *Nature* 256:495, 1975 or derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein or peptides derived therefrom over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, *Meth. Enzymol.* 70:419, 1980, the disclosure of which is incorporated herein by reference and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis *et al.* in *Basic Methods in Molecular Biology* Elsevier, New York. Section 21-2, the disclosure of which is incorporated herein by reference.

2. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein or peptides derived therefrom, which can be unmodified or modified to enhance

immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than others and may require the use of carriers and adjuvant. Also, host animals response vary depending on site of inoculations and doses, with both inadequate or
5 excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis. *et al*, *J. Clin. Endocrinol. Metab.* **33**:988-991 (1971), the disclosure of which is incorporated herein by reference..

Booster injections can be given at regular intervals, and antiserum harvested when
10 antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, *et al.*, Chap. 19 in: *Handbook of Experimental Immunology* D. Wier (ed) Blackwell (1973), the disclosure of which is incorporated herein by reference. Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 μ M).
15 Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: *Manual of Clinical Immunology*, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980), the disclosure of which is incorporated herein by reference..

Antibody preparations prepared according to either protocol are useful in
20 quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

25

V. Use of 5' ESTs or Sequences Obtainable Therefrom or Portions Thereof as Reagents

The 5' ESTs of the present invention (or cDNAs or genomic DNAs obtainable therefrom) may be used as reagents in isolation procedures, diagnostic assays, and
30 forensic procedures. For example, sequences from the 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) may be detectably labeled and used as probes to isolate

other sequences capable of hybridizing to them. In addition, sequences from 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) may be used to design PCR primers to be used in isolation, diagnostic, or forensic procedures.

5 1. Use of 5' ESTs or Sequences Obtainable Therefrom or Portions Thereof in Isolation, Diagnostic and Forensic Procedures

EXAMPLE 44

Preparation of PCR Primers and Amplification of DNA

10 The 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom) may be used to prepare PCR primers for a variety of applications, including isolation procedures for cloning nucleic acids capable of hybridizing to such sequences, diagnostic techniques and forensic techniques. The PCR primers are at least 10 bases, and preferably at least 12, 15, or 17 bases in length. More preferably, the PCR primers are at least 20-30 bases in length. In
15 some embodiments, the PCR primers may be more than 30 bases in length. It is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see Molecular Cloning to Genetic Engineering, White Ed. in *Methods in Molecular Biology* 67: Humana Press, Totowa 1997, the disclosure of which
20 is incorporated herein by reference. In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample.
25 The hybridized primers are extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites.

EXAMPLE 45Use of 5'ESTs as Probes

Probes derived from 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom), including full length cDNAs or genomic sequences, may be labeled with detectable labels familiar to those skilled in the art, including radioisotopes and non-radioactive labels, to provide a detectable probe. The detectable probe may be single stranded or double stranded and may be made using techniques known in the art, including *in vitro* transcription, nick translation, or kinase reactions. A nucleic acid sample containing a sequence capable of hybridizing to the labeled probe is contacted with the labeled probe. If the nucleic acid in the sample is double stranded, it may be denatured prior to contacting the probe. In some applications, the nucleic acid sample may be immobilized on a surface such as a nitrocellulose or nylon membrane. The nucleic acid sample may comprise nucleic acids obtained from a variety of sources, including genomic DNA, cDNA libraries, RNA, or tissue samples.

Procedures used to detect the presence of nucleic acids capable of hybridizing to the detectable probe include well known techniques such as Southern blotting, Northern blotting, dot blotting, colony hybridization, and plaque hybridization. In some applications, the nucleic acid capable of hybridizing to the labeled probe may be cloned into vectors such as expression vectors, sequencing vectors, or *in vitro* transcription vectors to facilitate the characterization and expression of the hybridizing nucleic acids in the sample. For example, such techniques may be used to isolate and clone sequences in a genomic library or cDNA library which are capable of hybridizing to the detectable probe as described in Example 30 above.

PCR primers made as described in Example 44 above may be used in forensic analyses, such as the DNA fingerprinting techniques described in Examples 46-50 below. Such analyses may utilize detectable probes or primers based on the sequences of the the 5' ESTs or of cDNAs or genomic DNAs isolated using the 5' ESTs.

EXAMPLE 46Forensic Matching by DNA Sequencing

In one exemplary method, DNA samples are isolated from forensic specimens of, for example, hair, semen, blood or skin cells by conventional methods. A panel of PCR primers based on a number of the 5' ESTs of Example 25, or cDNAs or genomic DNAs isolated

therefrom as described above, is then utilized in accordance with Example 44 to amplify DNA of approximately 100-200 bases in length from the forensic specimen. Corresponding sequences are obtained from a test subject. Each of these identification DNAs is then sequenced using standard techniques, and a simple database comparison determines the differences, if any, between the sequences from the subject and those from the sample. Statistically significant differences between the suspect's DNA sequences and those from the sample conclusively prove a lack of identity. This lack of identity can be proven, for example, with only one sequence. Identity, on the other hand, should be demonstrated with a large number of sequences, all matching. Preferably, a minimum of 50 statistically identical sequences of 100 bases in length are used to prove identity between the suspect and the sample.

EXAMPLE 47

Positive Identification by DNA Sequencing

The technique outlined in the previous example may also be used on a larger scale to provide a unique fingerprint-type identification of any individual. In this technique, primers are prepared from a large number of 5'EST sequences from Example 25, or cDNA or genomic DNA sequences obtainable therefrom. Preferably, 20 to 50 different primers are used. These primers are used to obtain a corresponding number of PCR-generated DNA segments from the individual in question in accordance with Example 44. Each of these DNA segments is sequenced, using the methods set forth in Example 46. The database of sequences generated through this procedure uniquely identifies the individual from whom the sequences were obtained. The same panel of primers may then be used at any later time to absolutely correlate tissue or other biological specimen with that individual.

25

EXAMPLE 48

Southern Blot Forensic Identification

The procedure of Example 47 is repeated to obtain a panel of at least 10 amplified sequences from an individual and a specimen. Preferably, the panel contains at least 50 amplified sequences. More preferably, the panel contains 100 amplified sequences. In some embodiments, the panel contains 200 amplified sequences. This PCR-generated DNA is then

30

digested with one or a combination of, preferably, four base specific restriction enzymes. Such enzymes are commercially available and known to those of skill in the art. After digestion, the resultant gene fragments are size separated in multiple duplicate wells on an agarose gel and transferred to nitrocellulose using Southern blotting techniques well known to those with skill in the art. For a review of Southern blotting see Davis *et al.* (Basic Methods in Molecular Biology, 1986, Elsevier Press. pp 62-65), the disclosure of which is incorporated herein by reference..

A panel of probes based on the sequences of 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom), or fragments thereof of at least 10 bases, are radioactively or colorimetrically labeled using methods known in the art, such as nick translation or end labeling, and hybridized to the Southern blot using techniques known in the art (Davis *et al.*, supra). Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the 5' EST (or cDNAs or genomic DNAs obtainable therefrom). More preferably, the probe comprises at least 20-30 consecutive nucleotides from the 5' EST (or cDNAs or genomic DNAs obtainable therefrom). In some embodiments, the probe comprises more than 30 nucleotides from the 5' EST (or cDNAs or genomic DNAs obtainable therefrom).

Preferably, at least 5 to 10 of these labeled probes are used, and more preferably at least about 20 or 30 are used to provide a unique pattern. The resultant bands appearing from the hybridization of a large sample of 5' EST (or cDNAs or genomic DNAs obtainable therefrom) will be a unique identifier. Since the restriction enzyme cleavage will be different for every individual, the band pattern on the Southern blot will also be unique. Increasing the number of 5' EST (or cDNAs or genomic DNAs obtainable therefrom) probes will provide a statistically higher level of confidence in the identification since there will be an increased number of sets of bands used for identification.

25

EXAMPLE 49

Dot Blot Identification Procedure

Another technique for identifying individuals using the 5' EST sequences disclosed herein utilizes a dot blot hybridization technique.

Genomic DNA is isolated from nuclei of subject to be identified. Oligonucleotide probes of approximately 30 bp in length are synthesized that correspond to at least 10,

30

preferably 50 sequences from the 5' ESTs or cDNAs or genomic DNAs obtainable therefrom. The probes are used to hybridize to the genomic DNA through conditions known to those in the art. The oligonucleotides are end labeled with P^{32} using polynucleotide kinase (Pharmacia). Dot Blots are created by spotting the genomic DNA onto nitrocellulose or the like using a vacuum dot blot manifold (BioRad, Richmond California). The nitrocellulose filter containing the genomic sequences is baked or UV linked to the filter, prehybridized and hybridized with labeled probe using techniques known in the art (Davis *et al.*, *supra*). The ^{32}P labeled DNA fragments are sequentially hybridized with successively stringent conditions to detect minimal differences between the 30 bp sequence and the DNA.

10 Tetramethylammonium chloride is useful for identifying clones containing small numbers of nucleotide mismatches (Wood *et al.*, *Proc. Natl. Acad. Sci. USA* 82(6):1585-1588, 1985) which is hereby incorporated by reference. A unique pattern of dots distinguishes one individual from another individual.

5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom) or oligonucleotides containing at least 10 consecutive bases from these sequences can be used as probes in the following alternative fingerprinting technique. Preferably, the probe comprises at least 12, 15, or 17 consecutive nucleotides from the 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom). More preferably, the probe comprises at least 20-30 consecutive nucleotides from the 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom). In some embodiments, the probe comprises more than 30 nucleotides from the 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom).

20

Preferably, a plurality of probes having sequences from different genes are used in the alternative fingerprinting technique. Example 50 below provides a representative alternative fingerprinting procedure in which the probes are derived from 5'EST.

25

EXAMPLE 50

Alternative "Fingerprint" Identification Technique

20-mer oligonucleotides are prepared from a large number, e.g. 50, 100, or 200, of 5'EST using commercially available oligonucleotide services such as Genset, Paris, France.

30 Cell samples from the test subject are processed for DNA using techniques well known to those with skill in the art. The nucleic acid is digested with restriction enzymes such as EcoRI

and XbaI: Following digestion, samples are applied to wells for electrophoresis. The procedure, as known in the art, may be modified to accommodate polyacrylamide electrophoresis, however in this example, samples containing 5 ug of DNA are loaded into wells and separated on 0.8% agarose gels. The gels are transferred onto nitrocellulose using
5 standard Southern blotting techniques.

10 ng of each of the oligonucleotides are pooled and end-labeled with ^{32}P . The nitrocellulose is prehybridized with blocking solution and hybridized with the labeled probes. Following hybridization and washing, the nitrocellulose filter is exposed to X-Omat AR X-ray film. The resulting hybridization pattern will be unique for each individual.

10 It is additionally contemplated within this example that the number of probe sequences used can be varied for additional accuracy or clarity.

The proteins encoded by the extended cDNAs may also be used to generate antibodies as explained in Examples 30 and 43 in order to identify the tissue type or cell
15 species from which a sample is derived as described in example 51.

EXAMPLE 51

Identification of Tissue Types or Cell Species by Means of Labeled Tissue Specific Antibodies

20 Identification of specific tissues is accomplished by the visualization of tissue specific antigens by means of antibody preparations according to Examples 30 and 43 which are conjugated, directly or indirectly to a detectable marker. Selected labeled antibody species bind to their specific antigen binding partner in tissue sections, cell suspensions, or in extracts of soluble proteins from a tissue sample to provide a pattern for qualitative or semi-qualitative
25 interpretation.

Antisera for these procedures must have a potency exceeding that of the native preparation, and for that reason, antibodies are concentrated to a mg/ml level by isolation of the gamma globulin fraction, for example, by ion-exchange chromatography or by ammonium sulfate fractionation. Also, to provide the most specific antisera, unwanted
30 antibodies, for example to common proteins, must be removed from the gamma globulin fraction, for example by means of insoluble immunoabsorbents, before the antibodies are

labeled with the marker. Either monoclonal or heterologous antisera is suitable for either procedure.

A. Immunohistochemical techniques

Purified, high-titer antibodies, prepared as described above, are conjugated to a detectable marker, as described, for example, by Fudenberg, Chap. 26 in: *Basic and Clinical Immunology*, 3rd Ed. Lange, Los Altos, California, 1980, or Rose, *et al.*, Chap. 12 in: *Methods in Immunodiagnosis*, 2d Ed. John Wiley and Sons, New York (1980), the disclosures of which are incorporated herein by reference.

A fluorescent marker, either fluorescein or rhodamine, is preferred, but antibodies can also be labeled with an enzyme that supports a color producing reaction with a substrate, such as horseradish peroxidase. Markers can be added to tissue-bound antibody in a second step, as described below. Alternatively, the specific antitissue antibodies can be labeled with ferritin or other electron dense particles, and localization of the ferritin coupled antigen-antibody complexes achieved by means of an electron microscope. In yet another approach, the antibodies are radiolabeled, with, for example ^{125}I , and detected by overlaying the antibody treated preparation with photographic emulsion.

Preparations to carry out the procedures can comprise monoclonal or polyclonal antibodies to a single protein or peptide identified as specific to a tissue type, for example, brain tissue, or antibody preparations to several antigenically distinct tissue specific antigens can be used in panels, independently or in mixtures, as required.

Tissue sections and cell suspensions are prepared for immunohistochemical examination according to common histological techniques. Multiple cryostat sections (about 4 μm , unfixed) of the unknown tissue and known control, are mounted and each slide covered with different dilutions of the antibody preparation. Sections of known and unknown tissues should also be treated with preparations to provide a positive control, a negative control, for example, pre-immune sera, and a control for non-specific staining, for example, buffer.

Treated sections are incubated in a humid chamber for 30 min at room temperature, rinsed, then washed in buffer for 30-45 min. Excess fluid is blotted away, and the marker developed.

If the tissue specific antibody was not labeled in the first incubation, it can be labeled at this time in a second antibody-antibody reaction, for example, by adding fluorescein- or enzyme-conjugated antibody against the immunoglobulin class of the antiserum-producing species, for example, fluorescein labeled antibody to mouse IgG. Such labeled sera are commercially available.

The antigen found in the tissues by the above procedure can be quantified by measuring the intensity of color or fluorescence on the tissue section, and calibrating that signal using appropriate standards.

B. Identification of tissue specific soluble proteins

The visualization of tissue specific proteins and identification of unknown tissues from that procedure is carried out using the labeled antibody reagents and detection strategy as described for immunohistochemistry; however the sample is prepared according to an electrophoretic technique to distribute the proteins extracted from the tissue in an orderly array on the basis of molecular weight for detection.

A tissue sample is homogenized using a Virtis apparatus; cell suspensions are disrupted by Dounce homogenization or osmotic lysis, using detergents in either case as required to disrupt cell membranes, as is the practice in the art. Insoluble cell components such as nuclei, microsomes, and membrane fragments are removed by ultracentrifugation, and the soluble protein-containing fraction concentrated if necessary and reserved for analysis.

A sample of the soluble protein solution is resolved into individual protein species by conventional SDS polyacrylamide electrophoresis as described, for example, by Davis, *et al.*, Section 19-2 in: *Basic Methods in Molecular Biology*, Leder ed., Elsevier, New York, 1986, the disclosure of which is incorporated herein by reference, using a range of amounts of polyacrylamide in a set of gels to resolve the entire molecular weight range of proteins to be detected in the sample. A size marker is run in parallel for purposes of estimating molecular weights of the constituent proteins. Sample size for analysis is a convenient volume of from 5 to 55 μ l, and containing from about 1 to 100 μ g protein. An aliquot of each of the resolved proteins is transferred by blotting to a nitrocellulose filter paper, a process that maintains the pattern of resolution. Multiple copies are prepared. The procedure, known as Western Blot Analysis, is well described in Davis, L. *et al.*, *supra* Section 19-3. One set of nitrocellulose blots is stained with Coomassie blue dye to visualize the entire set of proteins for comparison

with the antibody bound proteins. The remaining nitrocellulose filters are then incubated with a solution of one or more specific antisera to tissue specific proteins prepared as described in Examples 30 and 43. In this procedure, as in procedure A above, appropriate positive and negative sample and reagent controls are run.

5 In either procedure A or B, a detectable label can be attached to the primary tissue antigen-primary antibody complex according to various strategies and permutations thereof. In a straightforward approach, the primary specific antibody can be labeled; alternatively, the unlabeled complex can be bound by a labeled secondary anti-IgG antibody. In other approaches, either the primary or secondary antibody is conjugated to a biotin molecule, 10 which can, in a subsequent step, bind an avidin conjugated marker. According to yet another strategy, enzyme labeled or radioactive protein A, which has the property of binding to any IgG, is bound in a final step to either the primary or secondary antibody.

15 The visualization of tissue specific antigen binding at levels above those seen in control tissues to one or more tissue specific antibodies, prepared from the gene sequences identified from extended cDNA sequences, can identify tissues of unknown origin, for example, forensic samples, or differentiated tumor tissue that has metastasized to foreign bodily sites.

20 In addition to their applications in forensics and identification, 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) may be mapped to their chromosomal locations. Example 52 below describes radiation hybrid (RH) mapping of human chromosomal regions using 5'ESTs. Example 53 below describes a representative procedure for mapping an 5' EST to its location on a human chromosome. Example 54 below describes mapping of 5' ESTs on metaphase chromosomes by Fluorescence In Situ Hybridization (FISH). Those skilled in the art will appreciate that the method of 25 Examples 52-54 may also be used to map cDNAs or genomic DNAs obtainable from the 5' ESTs to their chromosomal locations.

2. Use of 5' ESTs or Sequences Obtainable Therefrom or Portions Thereof in Chromosome Mapping

EXAMPLE 52Radiation hybrid mapping of 5'ESTs to the human genome

Radiation hybrid (RH) mapping is a somatic cell genetic approach that can be used for high resolution mapping of the human genome. In this approach, cell lines containing one or more human chromosomes are lethally irradiated, breaking each chromosome into fragments whose size depends on the radiation dose. These fragments are rescued by fusion with cultured rodent cells, yielding subclones containing different portions of the human genome. This technique is described by Benham *et al.*, *Genomics* 4:509-517, 1989; and Cox *et al.*, *Science* 250:245-250, 1990, the entire contents of which are hereby incorporated by reference. The random and independent nature of the subclones permits efficient mapping of any human genome marker. Human DNA isolated from a panel of 80-100 cell lines provides a mapping reagent for ordering 5'EST. In this approach, the frequency of breakage between markers is used to measure distance, allowing construction of fine resolution maps as has been done using conventional ESTs (Schuler *et al.*, *Science* 274:540-546, 1996, hereby incorporated by reference).

RH mapping has been used to generate a high-resolution whole genome radiation hybrid map of human chromosome 17q22-q25.3 across the genes for growth hormone (GH) and thymidine kinase (TK) (Foster *et al.*, *Genomics* 33:185-192, 1996), the region surrounding the Gorlin syndrome gene (Obermayr *et al.*, *Eur. J. Hum. Genet.* 4:242-245, 1996), 60 loci covering the entire short arm of chromosome 12 (Raeymaekers *et al.*, *Genomics* 29:170-178, 1995), the region of human chromosome 22 containing the neurofibromatosis type 2 locus (Frazer *et al.*, *Genomics* 14:574-584, 1992) and 13 loci on the long arm of chromosome 5 (Warrington *et al.*, *Genomics* 11:701-708, 1991).

EXAMPLE 53Mapping of 5'ESTs to Human Chromosomes using PCR techniques

5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) may be assigned to human chromosomes using PCR based methodologies. In such approaches, oligonucleotide primer pairs are designed from the 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) to minimize the chance of amplifying through an intron. Preferably, the oligonucleotide primers are 18-23 bp in length and are designed for PCR amplification. The

creation of PCR primers from known sequences is well known to those with skill in the art. For a review of PCR technology see Erlich in *PCR Technology, Principles and Applications for DNA Amplification*, Freeman and Co., New York, 1992, the disclosure of which is incorporated herein by reference.

5 The primers are used in polymerase chain reactions (PCR) to amplify templates from total human genomic DNA. PCR conditions are as follows: 60 ng of genomic DNA is used as a template for PCR with 80 ng of each oligonucleotide primer, 0.6 unit of Taq polymerase, and 1 μ Cu of a 32 P-labeled deoxycytidine triphosphate. The PCR is performed in a microplate thermocycler (Techne) under the following conditions: 30 cycles of 94°C, 1.4 min;
10 55°C, 2 min; and 72°C, 2 min; with a final extension at 72°C for 10 min. The amplified products are analyzed on a 6% polyacrylamide sequencing gel and visualized by autoradiography. If the length of the resulting PCR product is identical to the distance between the ends of the primer sequences in the extended cDNA from which the primers are derived, then the PCR reaction is repeated with DNA templates from two panels of human-
15 rodent somatic cell hybrids, BIOS PCRable DNA (BIOS Corporation) and NIGMS Human-Rodent Somatic Cell Hybrid Mapping Panel Number 1 (NIGMS, Camden, NJ).

PCR is used to screen a series of somatic cell hybrid cell lines containing defined sets of human chromosomes for the presence of a given 5' EST (or cDNA or genomic DNA obtainable therefrom). DNA is isolated from the somatic hybrids and used as starting
20 templates for PCR reactions using the primer pairs from the 5' EST (or cDNA or genomic DNA obtainable therefrom). Only those somatic cell hybrids with chromosomes containing the human gene corresponding to the 5' EST (or cDNA or genomic DNA obtainable therefrom) will yield an amplified fragment. The 5' EST (or cDNA or genomic DNA obtainable therefrom) are assigned to a chromosome by analysis of the segregation pattern of
25 PCR products from the somatic hybrid DNA templates. The single human chromosome present in all cell hybrids that give rise to an amplified fragment is the chromosome containing that 5'EST (or cDNA or genomic DNA obtainable therefrom). For a review of techniques and analysis of results from somatic cell gene mapping experiments, see Ledbetter *et al.*, *Genomics* 6:475-481, 1990, the disclosure of which is incorporated herein by reference.

EXAMPLE 54

Mapping of Extended 5' ESTs to Chromosomes Using Fluorescence *In Situ*Hybridization

Fluorescence in situ hybridization allows the 5'EST (or cDNA or genomic DNA obtainable therefrom) to be mapped to a particular location on a given chromosome. The chromosomes to be used for fluorescence in situ hybridization techniques may be obtained from a variety of sources including cell cultures, tissues, or whole blood.

In a preferred embodiment, chromosomal localization of an 5'EST (or cDNA or genomic DNA obtainable therefrom) is obtained by FISH as described by Cherif *et al.* (*Proc. Natl. Acad. Sci. U.S.A.*, **87**:6639-6643, 1990), the disclosure of which is incorporated herein by reference. Metaphase chromosomes are prepared from phytohemagglutinin (PHA)-stimulated blood cell donors. PHA-stimulated lymphocytes from healthy males are cultured for 72 h in RPMI-1640 medium. For synchronization, methotrexate (10 μ M) is added for 17 h, followed by addition of 5-bromodeoxyuridine (5-BrdU, 0.1 mM) for 6 h. Colcemid (1 μ g/ml) is added for the last 15 min before harvesting the cells. Cells are collected, washed in RPMI, incubated with a hypotonic solution of KCl (75 mM) at 37°C for 15 min and fixed in three changes of methanol:acetic acid (3:1). The cell suspension is dropped onto a glass slide and air dried. The 5'EST (or cDNA or genomic DNA obtainable therefrom) is labeled with biotin-16 dUTP by nick translation according to the manufacturer's instructions (Bethesda Research Laboratories, Bethesda, MD), purified using a Sephadex G-50 column (Pharmacia, Upsala, Sweden) and precipitated. Just prior to hybridization, the DNA pellet is dissolved in hybridization buffer (50% formamide, 2 X SSC, 10% dextran sulfate, 1 mg/ml sonicated salmon sperm DNA, pH 7) and the probe is denatured at 70°C for 5-10 min.

Slides kept at -20°C are treated for 1 h at 37°C with RNase A (100 μ g/ml), rinsed three times in 2 X SSC and dehydrated in an ethanol series. Chromosome preparations are denatured in 70% formamide, 2 X SSC for 2 min at 70°C, then dehydrated at 4°C. The slides are treated with proteinase K (10 μ g/100 ml in 20 mM Tris-HCl, 2 mM CaCl₂) at 37°C for 8 min and dehydrated. The hybridization mixture containing the probe is placed on the slide, covered with a coverslip, sealed with rubber cement and incubated overnight in a humid chamber at 37°C. After hybridization and post-hybridization washes, the biotinylated probe is detected by avidin-FITC and amplified with additional layers of biotinylated goat anti-avidin

and avidin-FITC. For chromosomal localization, fluorescent R-bands are obtained as previously described (Cherif *et al.*, *supra.*). The slides are observed under a LEICA fluorescence microscope (DMRXA). Chromosomes are counterstained with propidium iodide and the fluorescent signal of the probe appears as two symmetrical yellow-green spots
5 on both chromatids of the fluorescent R-band chromosome (red). Thus, a particular 5'EST (or cDNA or genomic DNA obtainable therefrom) may be localized to a particular cytogenetic R-band on a given chromosome.

Once the 5'EST (or cDNA or genomic DNA obtainable therefrom) have been
10 assigned to particular chromosomes using the techniques described in Examples 52-54 above, they may be utilized to construct a high resolution map of the chromosomes on which they are located or to identify the chromosomes in a sample.

EXAMPLE 55

15 Use of 5'EST to Construct or Expand Chromosome Maps

Chromosome mapping involves assigning a given unique sequence to a particular chromosome as described above. Once the unique sequence has been mapped to a given chromosome, it is ordered relative to other unique sequences located on the same chromosome. One approach to chromosome mapping utilizes a series of yeast artificial
20 chromosomes (YACs) bearing several thousand long inserts derived from the chromosomes of the organism from which the extended cDNAs (or genomic DNAs obtainable therefrom) are obtained. This approach is described in Nagaraja *et al.*, *Genome Research* 7:210-222, 1997, the disclosure of which is incorporated herein by reference. Briefly, in this approach each chromosome is broken into overlapping pieces which are inserted into the YAC vector.
25 The YAC inserts are screened using PCR or other methods to determine whether they include the 5'EST (or cDNA or genomic DNA obtainable therefrom) whose position is to be determined. Once an insert has been found which includes the 5'EST (or cDNA or genomic DNA obtainable therefrom), the insert can be analyzed by PCR or other methods to determine whether the insert also contains other sequences known to be on the chromosome
30 or in the region from which the 5'EST (or cDNA or genomic DNA obtainable therefrom) was derived. This process can be repeated for each insert in the YAC library to determine the

location of each of the extended cDNAs (or genomic DNAs obtainable therefrom) relative to one another and to other known chromosomal markers. In this way, a high resolution map of the distribution of numerous unique markers along each of the organisms chromosomes may be obtained.

5

As described in Example 56 below extended cDNAs (or genomic DNAs obtainable therefrom) may also be used to identify genes associated with a particular phenotype, such as hereditary disease or drug response.

10 3. Use of 5'ESTs or Sequences Obtained Therefrom or Fragments Thereof in Gene Identification

EXAMPLE 56

Identification of genes associated with hereditary diseases or drug response

15 This example illustrates an approach useful for the association of 5'ESTs (or cDNA or genomic DNA obtainable therefrom) with particular phenotypic characteristics. In this example, a particular 5'EST (or cDNA or genomic DNA obtainable therefrom) is used as a test probe to associate that 5'EST (or cDNA or genomic DNA obtainable therefrom) with a particular phenotypic characteristic.

20 5'ESTs (or cDNA or genomic DNA obtainable therefrom) are mapped to a particular location on a human chromosome using techniques such as those described in Examples 52 and 53 or other techniques known in the art. A search of Mendelian Inheritance in Man (McKusick in *Mendelian Inheritance in Man* (available on line through Johns Hopkins University Welch Medical Library) reveals the region of the human chromosome which
25 contains the 5'EST (or cDNA or genomic DNA obtainable therefrom) to be a very gene rich region containing several known genes and several diseases or phenotypes for which genes have not been identified. The gene corresponding to this 5'EST (or cDNA or genomic DNA obtainable therefrom) thus becomes an immediate candidate for each of these genetic diseases.

30 Cells from patients with these diseases or phenotypes are isolated and expanded in culture. PCR primers from the 5'EST (or cDNA or genomic DNA obtainable

therefrom) are used to screen genomic DNA, mRNA or cDNA obtained from the patients. 5'ESTs (or cDNA or genomic DNA obtainable therefrom) that are not amplified in the patients can be positively associated with a particular disease by further analysis. Alternatively, the PCR analysis may yield fragments of different lengths when
5 the samples are derived from an individual having the phenotype associated with the disease than when the sample is derived from a healthy individual, indicating that the gene containing the 5'EST may be responsible for the genetic disease.

10 VI. Use of 5'EST (or cDNA or Genomic DNA Obtainable Therefrom) to Construct Vectors

The present 5'ESTs (or cDNA or genomic DNA obtainable therefrom) may also be used to construct secretion vectors capable of directing the secretion of the proteins encoded by genes therein. Such secretion vectors may facilitate the purification or enrichment of the proteins encoded by genes inserted therein by reducing the number of
15 background proteins from which the desired protein must be purified or enriched. Exemplary secretion vectors are described in Example 57 below.

1. Construction of Secretion Vectors

EXAMPLE 57

20 Construction of Secretion Vectors

The secretion vectors include a promoter capable of directing gene expression in the host cell, tissue, or organism of interest. Such promoters include the Rous Sarcoma Virus promoter, the SV40 promoter, the human cytomegalovirus promoter, and other promoters familiar to those skilled in the art.

25 A signal sequence from a 5' EST (or cDNAs or genomic DNAs obtainable therefrom) is operably linked to the promoter such that the mRNA transcribed from the promoter will direct the translation of the signal peptide. The host cell, tissue, or organism may be any cell, tissue, or organism which recognizes the signal peptide encoded by the signal sequence in the 5' EST (or cDNA or genomic DNA obtainable therefrom). Suitable hosts
30 include mammalian cells, tissues or organisms, avian cells, tissues, or organisms, insect cells, tissues or organisms, or yeast.

In addition, the secretion vector contains cloning sites for inserting genes encoding the proteins which are to be secreted. The cloning sites facilitate the cloning of the insert gene in frame with the signal sequence such that a fusion protein in which the signal peptide is fused to the protein encoded by the inserted gene is expressed from the mRNA transcribed from the promoter. The signal peptide directs the extracellular secretion of the fusion protein.

The secretion vector may be DNA or RNA and may integrate into the chromosome of the host, be stably maintained as an extrachromosomal replicon in the host, be an artificial chromosome, or be transiently present in the host. Many nucleic acid backbones suitable for use as secretion vectors are known to those skilled in the art, including retroviral vectors, SV40 vectors, Bovine Papilloma Virus vectors, yeast integrating plasmids, yeast episomal plasmids, yeast artificial chromosomes, human artificial chromosomes, P element vectors, baculovirus vectors, or bacterial plasmids capable of being transiently introduced into the host.

The secretion vector may also contain a polyA signal such that the polyA signal is located downstream of the gene inserted into the secretion vector.

After the gene encoding the protein for which secretion is desired is inserted into the secretion vector, the secretion vector is introduced into the host cell, tissue, or organism using calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection, viral particles or as naked DNA. The protein encoded by the inserted gene is then purified or enriched from the supernatant using conventional techniques such as ammonium sulfate precipitation, immunoprecipitation, immunochromatography, size exclusion chromatography, ion exchange chromatography, and HPLC. Alternatively, the secreted protein may be in a sufficiently enriched or pure state in the supernatant or growth media of the host to permit it to be used for its intended purpose without further enrichment.

The signal sequences may also be inserted into vectors designed for gene therapy. In such vectors, the signal sequence is operably linked to a promoter such that mRNA transcribed from the promoter encodes the signal peptide. A cloning site is located downstream of the signal sequence such that a gene encoding a protein whose secretion is desired may readily be inserted into the vector and fused to the signal sequence. The vector is introduced into an appropriate host cell. The protein expressed from the promoter is secreted extracellularly, thereby producing a therapeutic effect.

The 5' ESTs may also be used to clone sequences located upstream of the 5' ESTs which are capable of regulating gene expression, including promoter sequences, enhancer sequences, and other upstream sequences which influence transcription or translation levels. Once identified and cloned, these upstream regulatory sequences may be used in expression vectors designed to direct the expression of an inserted gene in a desired spatial, temporal, developmental, or quantitative fashion. Example 58 describes a method for cloning sequences upstream of the extended cDNAs or 5' ESTs.

2. Identification of Upstream Sequences With Promoting or Regulatory Activities

EXAMPLE 58

Use of Extended cDNAs or 5' ESTs to Clone Upstream Sequences from Genomic DNA

Sequences derived from extended cDNAs or 5' ESTs may be used to isolate the promoters of the corresponding genes using chromosome walking techniques. In one chromosome walking technique, which utilizes the GenomeWalker™ kit available from Clontech, five complete genomic DNA samples are each digested with a different restriction enzyme which has a 6 base recognition site and leaves a blunt end. Following digestion, oligonucleotide adapters are ligated to each end of the resulting genomic DNA fragments.

For each of the five genomic DNA libraries, a first PCR reaction is performed according to the manufacturer's instructions (which are incorporated herein by reference) using an outer adaptor primer provided in the kit and an outer gene specific primer. The gene specific primer should be selected to be specific for the extended cDNA or 5' EST of interest and should have a melting temperature, length, and location in the extended cDNA or 5'EST which is consistent with its use in PCR reactions. Each first PCR reaction contains 5 ng of genomic DNA, 5 µl of 10X Tth reaction buffer, 0.2 mM of each dNTP, 0.2 µM each of outer adaptor primer and outer gene specific primer, 1.1 mM of Mg(OAc)₂, and 1 µl of the Tth polymerase 50X mix in a total volume of 50 µl. The reaction cycle for the first PCR reaction is as follows: 1 min - 94°C / 2 sec - 94°C, 3 min - 72°C (7 cycles) / 2 sec - 94°C, 3 min - 67°C (32 cycles) / 5 min - 67°C.

The product of the first PCR reaction is diluted and used as a template for a second PCR reaction according to the manufacturer's instructions using a pair of nested

primers which are located internally on the amplicon resulting from the first PCR reaction. For example, 5 µl of the reaction product of the first PCR reaction mixture may be diluted 180 times. Reactions are made in a 50 µl volume having a composition identical to that of the first PCR reaction except the nested primers are used. The first
5 nested primer is specific for the adaptor, and is provided with the GenomeWalker™ kit. The second nested primer is specific for the particular extended cDNA or 5' EST for which the promoter is to be cloned and should have a melting temperature, length, and location in the extended cDNA or 5' EST which is consistent with its use in PCR reactions. The reaction parameters of the second PCR reaction are as follows: 1 min -
10 94°C / 2 sec - 94°C, 3 min - 72°C (6 cycles) / 2 sec - 94°C, 3 min - 67°C (25 cycles) / 5 min - 67°C. The product of the second PCR reaction is purified, cloned, and sequenced using standard techniques.

Alternatively, two or more human genomic DNA libraries can be constructed by using two or more restriction enzymes. The digested genomic DNA is cloned into vectors
15 which can be converted into single stranded, circular, or linear DNA. A biotinylated oligonucleotide comprising at least 15 nucleotides from the extended cDNA or 5' EST sequence is hybridized to the single stranded DNA. Hybrids between the biotinylated oligonucleotide and the single stranded DNA containing the extended cDNA or EST sequence are isolated as described in Example 29 above. Thereafter, the single stranded
20 DNA containing the extended cDNA or EST sequence is released from the beads and converted into double stranded DNA using a primer specific for the extended cDNA or 5' EST sequence or a primer corresponding to a sequence included in the cloning vector. The resulting double stranded DNA is transformed into bacteria. DNAs containing the 5' EST or extended cDNA sequences are identified by colony PCR or colony hybridization.

25

Once the upstream genomic sequences have been cloned and sequenced as described above, prospective promoters and transcription start sites within the upstream sequences may be identified by comparing the sequences upstream of the extended cDNAs or 5' ESTs with
30 databases containing known transcription start sites, transcription factor binding sites, or promoter sequences.

In addition, promoters in the upstream sequences may be identified using promoter reporter vectors as described in Example .

EXAMPLE 59

5 Identification of Promoters in Cloned Upstream Sequences

The genomic sequences upstream of the extended cDNAs or 5' ESTs are cloned into a suitable promoter reporter vector, such as the pSEAP-Basic, pSEAP-Enhancer, p β gal-Basic, p β gal-Enhancer, or pEGFP-1 Promoter Reporter vectors available from Clontech. Briefly, each of these promoter reporter vectors include multiple cloning sites positioned
10 upstream of a reporter gene encoding a readily assayable protein such as secreted alkaline phosphatase, β galactosidase, or green fluorescent protein. The sequences upstream of the extended cDNAs or 5' ESTs are inserted into the cloning sites upstream of the reporter gene in both orientations and introduced into an appropriate host cell. The level of reporter protein is assayed and compared to the level obtained from a vector which lacks an insert in the
15 cloning site. The presence of an elevated expression level in the vector containing the insert with respect to the control vector indicates the presence of a promoter in the insert. If necessary, the upstream sequences can be cloned into vectors which contain an enhancer for augmenting transcription levels from weak promoter sequences. A significant level of expression above that observed with the vector lacking an insert indicates that a promoter
20 sequence is present in the inserted upstream sequence.

Appropriate host cells for the promoter reporter vectors may be chosen based on the results of the above described determination of expression patterns of the extended cDNAs and ESTs. For example, if the expression pattern analysis indicates that the mRNA corresponding to a particular extended cDNA or 5' EST is expressed in fibroblasts, the
25 promoter reporter vector may be introduced into a human fibroblast cell line.

Promoter sequences within the upstream genomic DNA may be further defined by constructing nested deletions in the upstream DNA using conventional techniques such as Exonuclease III digestion. The resulting deletion fragments can be inserted into the promoter reporter vector to determine whether the deletion has reduced or obliterated promoter
30 activity. In this way, the boundaries of the promoters may be defined. If desired, potential individual regulatory sites within the promoter may be identified using site directed

mutagenesis or linker scanning to obliterate potential transcription factor binding sites within the promoter individually or in combination. The effects of these mutations on transcription levels may be determined by inserting the mutations into the cloning sites in the promoter reporter vectors.

5

EXAMPLE 60

Cloning and Identification of Promoters

Using the method described in Example 58 above with 5' ESTs, sequences upstream of several genes were obtained. Using the primer pairs GGG AAG ATG GAG ATA GTA
10 TTG CCT G (SEQ ID NO:29) and CTG CCA TGT ACA TGA TAG AGA GAT TC (SEQ ID NO:30), the promoter having the internal designation P13H2 (SEQ ID NO:31) was obtained.

Using the primer pairs GTA CCA GGGG ACT GTG ACC ATT GC (SEQ ID NO:32) and CTG TGA CCA TTG CTC CCA AGA GAG (SEQ ID NO:33), the promoter
15 having the internal designation P15B4 (SEQ ID NO:34) was obtained.

Using the primer pairs CTG GGA TGG AAG GCA CGG TA (SEQ ID NO:35) and GAG ACC ACA CAG CTA GAC AA (SEQ ID NO:36), the promoter having the internal designation P29B6 (SEQ ID NO:37) was obtained.

Figure 4 provides a schematic description of the promoters isolated and the way they
20 are assembled with the corresponding 5' tags. The upstream sequences were screened for the presence of motifs resembling transcription factor binding sites or known transcription start sites using the computer program MatInspector release 2.0, August 1996.

Table VII describes the transcription factor binding sites present in each of these promoters. The columns labeled matrix provides the name of the MatInspector matrix used.
25 The column labeled position provides the 5' position of the promoter site. Numeration of the sequence starts from the transcription site as determined by matching the genomic sequence with the 5' EST sequence. The column labeled "orientation" indicates the DNA strand on which the site is found, with the + strand being the coding strand as determined by matching the genomic sequence with the sequence of the 5' EST. The column labeled "score" provides
30 the MatInspector score found for this site. The column labeled "length" provides the length

of the site in nucleotides. The column labeled "sequence" provides the sequence of the site found.

Bacterial clones containing plasmids containing the promoter sequences described above described above are presently stored in the inventor's laboratories under the internal
5 identification numbers provided above. The inserts may be recovered from the deposited materials by growing an aliquot of the appropriate bacterial clone in the appropriate medium. The plasmid DNA can then be isolated using plasmid isolation procedures familiar to those skilled in the art such as alkaline lysis minipreps or large scale alkaline lysis plasmid isolation procedures. If desired the plasmid DNA may be further enriched by centrifugation on a
10 cesium chloride gradient, size exclusion chromatography, or anion exchange chromatography.

The plasmid DNA obtained using these procedures may then be manipulated using standard cloning techniques familiar to those skilled in the art. Alternatively, a PCR can be done with primers designed at both ends of the EST insertion. The PCR product which corresponds to the 5' EST can then be manipulated using standard cloning techniques familiar to those skilled
15 in the art.

The promoters and other regulatory sequences located upstream of the extended cDNAs or 5' ESTs may be used to design expression vectors capable of directing the expression of an inserted gene in a desired spatial, temporal, developmental, or quantitative manner. A promoter capable of directing the desired spatial, temporal, developmental, and
20 quantitative patterns may be selected using the results of the expression analysis described in Example 26 above. For example, if a promoter which confers a high level of expression in muscle is desired, the promoter sequence upstream of an extended cDNA or 5' EST derived from an mRNA which is expressed at a high level in muscle, as determined by the method of Example 26, may be used in the expression vector.

25 Preferably, the desired promoter is placed near multiple restriction sites to facilitate the cloning of the desired insert downstream of the promoter, such that the promoter is able to drive expression of the inserted gene. The promoter may be inserted in conventional nucleic acid backbones designed for extrachromosomal replication, integration into the host chromosomes or transient expression. Suitable backbones for the present expression vectors
30 include retroviral backbones, backbones from eukaryotic episomes such as SV40 or Bovine Papilloma Virus, backbones from bacterial episomes, or artificial chromosomes.

Preferably, the expression vectors also include a polyA signal downstream of the multiple restriction sites for directing the polyadenylation of mRNA transcribed from the gene inserted into the expression vector.

Following the identification of promoter sequences using the procedures of Examples 58-60, proteins which interact with the promoter may be identified as described in Example 61 below.

EXAMPLE 61

Identification of Proteins Which Interact with Promoter Sequences, Upstream

Regulatory Sequences, or mRNA

Sequences within the promoter region which are likely to bind transcription factors may be identified by homology to known transcription factor binding sites or through conventional mutagenesis or deletion analyses of reporter plasmids containing the promoter sequence. For example, deletions may be made in a reporter plasmid containing the promoter sequence of interest operably linked to an assayable reporter gene. The reporter plasmids carrying various deletions within the promoter region are transfected into an appropriate host cell and the effects of the deletions on expression levels is assessed. Transcription factor binding sites within the regions in which deletions reduce expression levels may be further localized using site directed mutagenesis, linker scanning analysis, or other techniques familiar to those skilled in the art.

Nucleic acids encoding proteins which interact with sequences in the promoter may be identified using one-hybrid systems such as those described in the manual accompanying the Matchmaker One-Hybrid System kit available from Clontech (Catalog No. K1603-1), the disclosure of which is incorporated herein by reference. Briefly, the Matchmaker One-hybrid system is used as follows. The target sequence for which it is desired to identify binding proteins is cloned upstream of a selectable reporter gene and integrated into the yeast genome. Preferably, multiple copies of the target sequences are inserted into the reporter plasmid in tandem. A library comprised of fusions between cDNAs to be evaluated for the ability to bind to the promoter and the activation domain of a yeast transcription factor, such as GAL4, is transformed into the yeast strain containing the integrated reporter sequence. The yeast are plated on selective media to

select cells expressing the selectable marker linked to the promoter sequence. The colonies which grow on the selective media contain genes encoding proteins which bind the target sequence. The inserts in the genes encoding the fusion proteins are further characterized by sequencing. In addition, the inserts may be inserted into expression
5 vectors or *in vitro* transcription vectors. Binding of the polypeptides encoded by the inserts to the promoter DNA may be confirmed by techniques familiar to those skilled in the art, such as gel shift analysis or DNase protection analysis.

10 VII. Use of 5' ESTs (or cDNAs or Genomic DNAs Obtainable Therefrom) in Gene Therapy

The present invention also comprises the use of 5'ESTs (or cDNA or genomic DNA obtainable therefrom) in gene therapy strategies, including antisense and triple helix strategies as described in Examples 62 and 63 below. In antisense approaches, nucleic acid sequences complementary to an mRNA are hybridized to the mRNA intracellularly, thereby blocking the
15 expression of the protein encoded by the mRNA. The antisense sequences may prevent gene expression through a variety of mechanisms. For example, the antisense sequences may inhibit the ability of ribosomes to translate the mRNA. Alternatively, the antisense sequences may block transport of the mRNA from the nucleus to the cytoplasm, thereby limiting the amount of mRNA available for translation. Another mechanism through which antisense
20 sequences may inhibit gene expression is by interfering with mRNA splicing. In yet another strategy, the antisense nucleic acid may be incorporated in a ribozyme capable of specifically cleaving the target mRNA.

EXAMPLE 62

25 Preparation and Use of Antisense Oligonucleotides

The antisense nucleic acid molecules to be used in gene therapy may be either DNA or RNA sequences. They may comprise a sequence complementary to the sequence of the 5'EST (or cDNA or genomic DNA obtainable therefrom). The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an intracellular
30 duplex with sufficient stability to inhibit the expression of the mRNA in the duplex. Strategies for designing antisense nucleic acids suitable for use in gene therapy are disclosed in Green *et*

al., *Ann. Rev. Biochem.* 55:569-597, 1986; and Izant and Weintraub, *Cell* 36:1007-1015, 1984, which are hereby incorporated by reference.

In some strategies, antisense molecules are obtained from a nucleotide sequence encoding a protein by reversing the orientation of the coding region with respect to a promoter so as to transcribe the opposite strand from that which is normally transcribed in the cell. The antisense molecules may be transcribed using *in vitro* transcription systems such as those which employ T7 or SP6 polymerase to generate the transcript. Another approach involves transcription of the antisense nucleic acids *in vivo* by operably linking DNA containing the antisense sequence to a promoter in an expression vector.

Alternatively, oligonucleotides which are complementary to the strand normally transcribed in the cell may be synthesized *in vitro*. Thus, the antisense nucleic acids are complementary to the corresponding mRNA and are capable of hybridizing to the mRNA to create a duplex. In some embodiments, the antisense sequences may contain modified sugar phosphate backbones to increase stability and make them less sensitive to RNase activity.

Examples of modifications suitable for use in antisense strategies are described by Rossi *et al.*, *Pharmacol. Ther.* 50(2):245-254, 1991, which is hereby incorporated by reference.

Various types of antisense oligonucleotides complementary to the sequence of the 5'EST (or cDNA or genomic DNA obtainable therefrom) may be used. In one preferred embodiment, stable and semi-stable antisense oligonucleotides described in International Application No. PCT WO94/23026, hereby incorporated by reference, are used. In these molecules, the 3' end or both the 3' and 5' ends are engaged in intramolecular hydrogen bonding between complementary base pairs. These molecules are better able to withstand exonuclease attacks and exhibit increased stability compared to conventional antisense oligonucleotides.

In another preferred embodiment, the antisense oligodeoxynucleotides against herpes simplex virus types 1 and 2 described in International Application No. WO 95/04141, hereby incorporated by reference, are used.

In yet another preferred embodiment, the covalently cross-linked antisense oligonucleotides described in International Application No. WO 96/31523, hereby incorporated by reference, are used. These double- or single-stranded oligonucleotides comprise one or more, respectively, inter- or intra-oligonucleotide covalent cross-linkages,

wherein the linkage consists of an amide bond between a primary amine group of one strand and a carboxyl group of the other strand or of the same strand, respectively, the primary amine group being directly substituted in the 2' position of the strand nucleotide monosaccharide ring, and the carboxyl group being carried by an aliphatic spacer group substituted on a nucleotide or nucleotide analog of the other strand or the same strand, respectively.

The antisense oligodeoxynucleotides and oligonucleotides disclosed in International Application No. WO 92/18522, incorporated by reference, may also be used. These molecules are stable to degradation and contain at least one transcription control recognition sequence which binds to control proteins and are effective as decoys therefore. These molecules may contain "hairpin" structures, "dumbbell" structures, "modified dumbbell" structures, "cross-linked" decoy structures and "loop" structures.

In another preferred embodiment, the cyclic double-stranded oligonucleotides described in European Patent Application No. 0 572 287 A2, hereby incorporated by reference are used. These ligated oligonucleotide "dumbbells" contain the binding site for a transcription factor and inhibit expression of the gene under control of the transcription factor by sequestering the factor.

Use of the closed antisense oligonucleotides disclosed in International Application No. WO 92/19732, hereby incorporated by reference, is also contemplated. Because these molecules have no free ends, they are more resistant to degradation by exonucleases than are conventional oligonucleotides. These oligonucleotides may be multifunctional, interacting with several regions which are not adjacent to the target mRNA.

The appropriate level of antisense nucleic acids required to inhibit gene expression may be determined using *in vitro* expression analysis. The antisense molecule may be introduced into the cells by diffusion, injection, infection, transfection or h-region-mediated import using procedures known in the art. For example, the antisense nucleic acids can be introduced into the body as a bare or naked oligonucleotide, oligonucleotide encapsulated in lipid, oligonucleotide sequence encapsulated by viral protein, or as an oligonucleotide operably linked to a promoter contained in an expression vector. The expression vector may be any of a variety of expression vectors known in the art, including retroviral or viral vectors,

vectors capable of extrachromosomal replication, or integrating vectors. The vectors may be DNA or RNA.

The antisense molecules are introduced onto cell samples at a number of different concentrations preferably between 1×10^{-10} M to 1×10^{-4} M. Once the minimum concentration that can adequately control gene expression is identified, the optimized dose is translated into a dosage suitable for use *in vivo*. For example, an inhibiting concentration in culture of 1×10^{-7} translates into a dose of approximately 0.6 mg/kg bodyweight. Levels of oligonucleotide approaching 100 mg/kg bodyweight or higher may be possible after testing the toxicity of the oligonucleotide in laboratory animals. It is additionally contemplated that cells from the vertebrate are removed, treated with the antisense oligonucleotide, and reintroduced into the vertebrate.

It is further contemplated that the antisense oligonucleotide sequence is incorporated into a ribozyme sequence to enable the antisense to specifically bind and cleave its target mRNA. For technical applications of ribozyme and antisense oligonucleotides see Rossi *et al.*, *supra*.

In a preferred application of this invention, the polypeptide encoded by the gene is first identified, so that the effectiveness of antisense inhibition on translation can be monitored using techniques that include but are not limited to antibody-mediated tests such as RIAs and ELISA, functional assays, or radiolabeling.

The 5' ESTs of the present invention (or cDNAs or genomic DNAs obtainable therefrom) may also be used in gene therapy approaches based on intracellular triple helix formation. Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for studying alterations in cell activity as it is associated with a particular gene. The 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom) of the present invention or, more preferably, a portion of those sequences, can be used to inhibit gene expression in individuals having diseases associated with expression of a particular gene. Similarly, a portion of 5' EST sequences (or cDNAs or genomic DNAs obtainable therefrom) can be used to study the effect of inhibiting transcription of a particular gene within a cell. Traditionally, homopurine sequences were considered the most useful for triple helix strategies. However, homopyrimidine sequences can also inhibit gene expression. Such homopyrimidine oligonucleotides bind to the major groove at

homopurine:homopyrimidine sequences. Thus, both types of sequences from the 5'EST or from the gene corresponding to the 5'EST are contemplated within the scope of this invention.

5

EXAMPLE 63

Preparation and Use of Triple Helix Probes

The sequences of the 5' ESTs (or cDNAs or genomic DNAs obtainable therefrom) are scanned to identify 10-mer to 20-mer homopyrimidine or homopurine stretches which could be used in triple-helix based strategies for inhibiting gene expression. Following
10 identification of candidate homopyrimidine or homopurine stretches, their efficiency in inhibiting gene expression is assessed by introducing varying amounts of oligonucleotides containing the candidate sequences into tissue culture cells which normally express the target gene. The oligonucleotides may be prepared on an oligonucleotide synthesizer or they may be purchased commercially from a company specializing in custom oligonucleotide synthesis,
15 such as GENSET, Paris, France.

The oligonucleotides may be introduced into the cells using a variety of methods known to those skilled in the art, including but not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake.

Treated cells are monitored for altered cell function or reduced gene expression using
20 techniques such as Northern blotting, RNase protection assays, or PCR based strategies to monitor the transcription levels of the target gene in cells which have been treated with the oligonucleotide. The cell functions to be monitored are predicted based upon the homologies of the target gene corresponding to the extended cDNA from which the oligonucleotide was derived with known gene sequences that have been associated with a particular function. The
25 cell functions can also be predicted based on the presence of abnormal physiologies within cells derived from individuals with a particular inherited disease, particularly when the extended cDNA is associated with the disease using techniques described in Example 56.

The oligonucleotides which are effective in inhibiting gene expression in tissue culture cells may then be introduced *in vivo* using the techniques described above and in Example 62
30 at a dosage calculated based on the *in vitro* results, as described in Example 62.

In some embodiments, the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3' end of the alpha oligonucleotide to stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin *et al.*, *Science* 245:967-971, 1989, which is hereby incorporated by this reference.

EXAMPLE 64

Use of cDNAs Obtained Using the 5' ESTs to Express an Encoded Protein in a Host

10

Organism

The cDNAs obtained as described above using the 5' ESTs of the present invention may also be used to express an encoded protein in a host organism to produce a beneficial effect. In such procedures, the encoded protein may be transiently expressed in the host organism or stably expressed in the host organism. The encoded protein may have any of the activities described above. The encoded protein may be a protein which the host organism lacks or, alternatively, the encoded protein may augment the existing levels of the protein in the host organism.

A full length extended cDNA encoding the signal peptide and the mature protein, or an extended cDNA encoding only the mature protein is introduced into the host organism. The extended cDNA may be introduced into the host organism using a variety of techniques known to those of skill in the art. For example, the extended cDNA may be injected into the host organism as naked DNA such that the encoded protein is expressed in the host organism, thereby producing a beneficial effect.

Alternatively, the extended cDNA may be cloned into an expression vector downstream of a promoter which is active in the host organism. The expression vector may be any of the expression vectors designed for use in gene therapy, including viral or retroviral vectors. The expression vector may be directly introduced into the host organism such that the encoded protein is expressed in the host organism to produce a beneficial effect. In another approach, the expression vector may be introduced into cells *in vitro*. Cells containing the expression vector are thereafter selected and introduced into the host organism, where they express the encoded protein to produce a beneficial effect.

EXAMPLE 65

Use of Signal Peptides Encoded by 5' ESTs or Sequences obtained Therefrom
to Import Proteins Into Cells

The short core hydrophobic region (h) of signal peptides encoded by the 5'ESTS or
5 extended cDNAs derived from SEQ ID NOs: 38-291 may also be used as a carrier to import
a peptide or a protein of interest, so-called cargo, into tissue culture cells (Lin *et al.*, *J. Biol.
Chem.*, **270**: 14225-14258, 1995; Du *et al.*, *J. Peptide Res.*, **51**: 235-243, 1998; Rojas *et al.*,
Nature Biotech., **16**: 370-375, 1998).

When cell permeable peptides of limited size (approximately up to 25 amino acids)
10 are to be translocated across cell membrane, chemical synthesis may be used in order to add
the h region to either the C-terminus or the N-terminus to the cargo peptide of interest.
Alternatively, when longer peptides or proteins are to be imported into cells, nucleic acids can
be genetically engineered, using techniques familiar to those skilled in the art, in order to link
the extended cDNA sequence encoding the h region to the 5' or the 3' end of a DNA
15 sequence coding for a cargo polypeptide. Such genetically engineered nucleic acids are then
translated either *in vitro* or *in vivo* after transfection into appropriate cells, using conventional
techniques to produce the resulting cell permeable polypeptide. Suitable hosts cells are then
simply incubated with the cell permeable polypeptide which is then translocated across the
membrane.

20 This method may be applied to study diverse intracellular functions and cellular
processes. For instance, it has been used to probe functionally relevant domains of
intracellular proteins and to examine protein-protein interactions involved in signal
transduction pathways (Lin *et al.*, *supra*; Lin *et al.*, *J. Biol. Chem.*, **271**: 5305-5308, 1996;
Rojas *et al.*, *J. Biol. Chem.*, **271**: 27456-27461, 1996; Liu *et al.*, *Proc. Natl. Acad. Sci. USA*,
25 **93**: 11819-11824, 1996; Rojas *et al.*, *Bioch. Biophys. Res. Commun.*, **234**: 675-680, 1997).

Such techniques may be used in cellular therapy to import proteins producing
therapeutic effects. For instance, cells isolated from a patient may be treated with imported
therapeutic proteins and then re-introduced into the host organism.

Alternatively, the h region of signal peptides of the present invention could be used in
30 combination with a nuclear localization signal to deliver nucleic acids into cell nucleus. Such
oligonucleotides may be antisense oligonucleotides or oligonucleotides designed to form

triple helixes, as described in examples 62 and 63 respectively, in order to inhibit processing and/or maturation of a target cellular RNA.

As discussed above, the cDNAs or portions thereof obtained using the 5' ESTs of the present invention can be used for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on Southern gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination for expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris *et al.*, *Cell* 75:791-803, 1993, the disclosure of which is hereby incorporated by reference) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The proteins or polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the other protein with which binding occurs or to identify inhibitors of the binding interaction. Proteins

involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

5 Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation *Molecular Cloning; A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, Fritsch and Maniatis eds., 1989, and *Methods in Enzymology; Guide to Molecular Cloning Techniques*, Academic Press, Berger and Kimmel eds., 1987.

10 Polynucleotides and proteins of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the protein or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid
15 preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the protein or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

 Although this invention has been described in terms of certain preferred embodiments, other embodiments which will be apparent to those of ordinary skill in the art in view of the
20 disclosure herein are also within the scope of this invention. Accordingly, the scope of the invention is intended to be defined only by reference to the appended claims. All documents cited herein are incorporated herein by reference in their entirety.

Step	Search characteristic		Selection Characteristics		
	Program	Strand	Parameters	Identity (%)	Length (bp)
Miscellaneous	blastn	both	S=61 X=16	90	17
tRNA	fasta	both	-	80	60
rRNA	blastn	both	S=108	80	40
mtRNA	blastn	both	S=108	80	40
Prokaryotic	blastn	both	S=144	90	40
Fungal	blastn	both	S=144	90	40
Alu	fasta*	both	-	70	40
L1	blastn	both	S=72	70	40
Repeats	blastn	both	S=72	70	40
Promoters	blastn	top	S=54 X=16	90	15†
Vertebrate	fasta*	both	S=108	90	30
ESTs	blastn	both	S=108 X=16	90	30
Proteins	blastx [‡]	top	E = 0.001	-	-

Table 1: Parameters used for each step of EST analysis

* use "Quick Fast" Database scanner

† alignment further constrained to begin closer than 10bp to EST's end

‡ using BLOSUM62 substitution matrix

TABLE II

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HEIJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID38	new	15	Liver	22-6-1-A10-PU
ID39	new	13.2	Fetal liver	
			Ovary	77-16-3-B7-PU
			Hypertrophic prostate	
ID40	new	13.1	Brain	
			Fetal brain	47-47-1-F2-PU
			Substantia nigra	
ID41	new	11.6	Fetal kidney	58-12-2-E11-PU
			Cancerous prostate	
ID42	new	10.7	Liver	21-4-2-D1-PU
			Kidney	
ID43	new	9.6	Hypertrophic prostate	77-38-4-B2-PU
			Cancerous prostate	
			Large intestine	
ID44	new	9.4	Fetal kidney	76-10-2-B7-PU
			Cancerous prostate	
ID45	new	9.4	Prostate	33-99-2-G8-PU
			Brain	
ID46	new	9.1	Hypertrophic prostate	78-32-2-C2-PU
			Normal prostate	
			Brain	
ID47	new	9.1	Ovary	26-40-3-D6-PU
			Brain	
ID48	new	8	Fetal kidney	33-106-2-F10-PU
			Brain	
ID49	new	7.8	Fetal kidney	58-38-1-A2-PU
			Lung (cells)	
ID50	new	7.4	Lymph ganglia	62-10-3-A11-PU
			Surrenals	
ID51	new	7.4	Hypertrophic prostate	76-45-1-F5-PU
			Cancerous prostate	
ID52	new	7.1	Fetal kidney	37-10-3-D7-PU
			Lung (cells)	
			Umbilical cord	
			Hypertrophic prostate	
			Cancerous prostate	
			Substantia nigra	
ID53	new	6.9	Hypertrophic prostate	78-16-2-B12-PU
			Normal prostate	
			Lymph ganglia	
			Spleen	
ID54	new	6.8	Fetal brain	33-38-2-A4-PU
			Brain	
ID55	new	6.7	Heart	47-25-4-A2-PU
			Spleen	
			Substantia nigra	
ID56	new	6.3	Fetal brain	20-10-3-D9-PU
			Spleen	
ID57	new	6.3	Hypertrophic prostate	84-5-1-C9-PU

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HELNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID58	new	6.3	Thyroid Prostate Hypertrophic prostate Normal prostate	76-40-1-A8-PU
ID59	new	6.3	Cancerous prostate Fetal kidney Normal prostate Hypertrophic prostate Cancerous prostate	76-5-1-F4-PU
ID60	new	6.3	Fetal kidney Hypertrophic prostate Kidney	77-25-3-H5-PU
ID61	new	5.7	Prostate Lymph ganglia Lung	42-1-4-H1-PU
ID62	new	5.6	Brain Lymph ganglia Pancreas	33-80-4-E4-PU
ID63	new	5.6	Fetal kidney Normal prostate	58-47-2-E11-PU
ID64	new	5.6	Muscle Brain	33-56-4-F4-PU
ID65	new	5.5	Placenta Lung (cells) Colon Cancerous prostate Normal prostate	23-1-4-F6-PU
ID66	new	5.3	Cancerous prostate Normal prostate	76-44-2-F7-PU
ID67	new	5.2	Cancerous prostate Hypertrophic prostate Cancerous prostate	76-19-1-E9-PU
ID68	new	5.1	Colon Normal prostate Kidney	78-31-1-D12-PU
ID69	new	4.9	Prostate Spleen	20-1-4-H6-PU
ID70	new	4.9	Lymphocytes Cancerous prostate	24-3-4-C4-PU
ID71	new	4.7	Kidney Brain	33-102-2-C9-PU
ID72	new	4.7	Colon Lymph ganglia	48-47-3-A5-PU
ID73	new	4.6	Placenta Hypertrophic prostate	77-2-3-D1-PU
ID74	new	4.6	Normal prostate Thyroid Cancerous prostate Substantia nigra	76-3-3-C7-PU
ID75	new	4.5	Fetal kidney Large intestine	83-1-3-H6-PU
ID76	new	4.4	Fetal brain Brain	33-7-2-D11-PU

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HEIJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID77	new	4	Normal prostate	78-28-2-G12-PU
ID78	new	3.9	Substantia nigra	
ID79	new	3.9	Normal prostate	76-23-3-D8-PU
ID80	new	3.8	Cancerous prostate	
ID81	new	3.8	Heart	48-3-3-H9-PU
ID82	new	3.8	Lymph ganglia	
ID83	new	3.7	Brain	42-2-4-B8-PU
ID84	new	3.5	Lung	
ID85	new	3.5	Normal prostate	77-37-2-H1-PU
ID86	ext-est-not-vrt	13.4	Hypertrophic prostate	
ID87	ext-est-not-vrt	12.6	Lung (cells)	51-37-4-B1-PU
ID88	ext-est-not-vrt	11.8	Testis	
ID89	ext-est-not-vrt	11.2	Lung	
ID90	ext-est-not-vrt	10.3	Ovary	23-9-4-G9-PU
ID91	ext-est-not-vrt	10.1	Lung (cells)	
ID92	ext-est-not-vrt	9.8	Colon	
ID93	ext-est-not-vrt	9.3	Normal prostate	
ID94	ext-est-not-vrt	8.4	Ovary	27-3-2-B6-PU
ID95	ext-est-not-vrt	7.8	Muscle	
ID96	ext-est-not-vrt	7.4	Hypertrophic prostate	
			Normal prostate	76-30-3-B7-PU
			Hypertrophic prostate	
			Cancerous prostate	
			Ovary	76-9-4-G9-PU
			Prostate	
			Cancerous prostate	
			Normal prostate	78-25-4-H1-PU
			Hypertrophic prostate	
			Fetal kidney	77-1-4-D10-PU
			Hypertrophic prostate	
			Lung (cells)	78-37-1-A12-PU
			Normal prostate	
			Cancerous prostate	
			Umbilical cord	37-10-2-C10-PU
			Hypertrophic prostate	
			Brain	76-16-1-H5-PU
			Cancerous prostate	
			Lymphocytes	24-1-4-G11-PU
			Lung (cells)	
			Umbilical cord	
			Normal prostate	
			Thyroid	48-51-2-C10-PU
			Heart	
			Lymph ganglia	
			Lung	
				33-97-4-G8-PU
			Fetal brain	33-22-1-F9-PU
			Brain	
			Ovary	37-7-4-E7-PU
			Liver	
			Umbilical cord	

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HELINE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID97	ext-est-not-vrt	7.2	Kidney Surrenals Muscle Liver Dystrophic muscle Normal prostate Testis Cancerous prostate Lymph ganglia Large intestine	27-12-3-H8-PU
ID98	ext-est-not-vrt	7.1	Fetal kidney Ovary	58-23-4-G9-PU
ID99	ext-est-not-vrt	6.9	Placenta	58-34-2-H8-PU
ID100	ext-est-not-vrt	6.7	Fetal kidney Fetal kidney Fetal brain Umbilical cord Heart	37-9-1-D4-PU
ID101	ext-est-not-vrt	6.6	Fetal liver Fetal kidney Liver Thyroid Kidney Cancerous prostate Lung (cells) Normal prostate Lymph ganglia	58-5-3-A8-PU
ID102	ext-est-not-vrt	6.6	Cancerous prostate Normal prostate	76-35-1-A11-PU
ID103	ext-est-not-vrt	5.4	Hypertrophic prostate Lung (cells)	77-35-2-E10-PU
ID104	ext-est-not-vrt	5.4	Fetal kidney Fetal brain Normal prostate	58-52-4-D8-PU
ID105	ext-est-not-vrt	5.3	Cancerous prostate Substantia nigra	47-26-3-D2-PU
ID106	ext-est-not-vrt	5.1	Cancerous prostate Fetal brain Lung (cells) Brain	30-9-1-G8-PU
ID107	ext-est-not-vrt	4.9	Lung Brain	33-98-1-C6-PU
ID108	ext-est-not-vrt	4.5	Ovary Prostate Normal prostate Brain	78-26-1-B12-PU
ID109	ext-est-not-vrt	4.2	Fetal kidney Cancerous prostate Normal prostate	58-7-2-F8-PU
ID110	ext-est-not-vrt	3.7	Fetal kidney Ovary	58-33-1-F9-PU

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HEIJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID111	ext-est-not-vrt	3.6	Prostate Normal prostate	33-19-1-F1-PU
ID112	ext-est-not-vrt	3.5	Brain Lymph ganglia Fetal kidney	
ID113	ext-est-not-vrt	3.5	Liver Kidney Brain Ovary	58-14-2-D3-PU
ID114	est-not-ext	13.9	Hypertrophic prostate Fetal kidney	26-40-2-B2-PU
ID115	est-not-ext	13.9	Cancerous prostate Normal prostate Fetal kidney	58-52-4-F10-PU
ID116	est-not-ext	11.6	Lung (cells) Ovary	58-15-1-H6-PU
ID117	est-not-ext	11.6	Dystrophic muscle Cancerous prostate Uterus Testis Lymph ganglia Surrenals	51-29-2-B2-PU
ID118	est-not-ext	11.6	Lymph ganglia Large intestine Umbilical cord	48-7-1-F2-PU
ID119	est-not-ext	11.4	Pancreas Heart	37-6-1-E12-PU
ID120	est-not-ext	11.2	Brain Dystrophic muscle	67-3-4-G7-PU
ID121	est-not-ext	11	Brain Ovary Heart Kidney	33-35-4-F4-PU
ID122	est-not-ext	10.5	Cancerous prostate Lymph ganglia Lung	48-14-1-A11-PU
ID123	est-not-ext	10	Umbilical cord Normal prostate Fetal kidney	37-11-1-G2-PU
ID124	est-not-ext	9.5	Cancerous prostate Normal prostate Brain Fetal kidney	58-3-4-G2-PU
ID125	est-not-ext	9.5	Cancerous prostate Umbilical cord Normal prostate Placenta	76-18-1-F6-PU
ID126	est-not-ext	9.3	Muscle Substantia nigra Ovary Cancerous prostate	47-24-2-C1-PU

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HEIJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
			Umbilical cord	
			Colon	
			Normal prostate	
			Testis	
ID127	est-not-ext	9.3	Cancerous prostate	47-37-2-E3-PU
			Normal prostate	
ID128	est-not-ext	9.3	Substantia nigra	
			Spleen	27-16-1-E4-PU
			Muscle	
ID129	est-not-ext	9.3	Colon	47-5-1-G3-PU
			Substantia nigra	
ID130	est-not-ext	9.2	Ovary	57-2-4-E11-PU
			Hypertrophic prostate	
			Fetal brain	
ID131	est-not-ext	9	Cancerous prostate	76-32-1-G12-PU
			Normal prostate	
ID132	est-not-ext	8.9	Fetal kidney	77-25-1-C6-PU
			Hypertrophic prostate	
			Placenta	
			Normal prostate	
			Brain	
ID133	est-not-ext	8.8	Dystrophic muscle	37-7-2-B11-PU
			Umbilical cord	
			Brain	
ID134	est-not-ext	8.8	Fetal kidney	77-7-3-C8-PU
			Dystrophic muscle	
			Hypertrophic prostate	
			Thyroid	
			Cancerous prostate	
			Fetal brain	
			Muscle	
			Lung (cells)	
			Normal prostate	
			Brain	
			Lymph ganglia	
			Large intestine	
ID135	est-not-ext	8.7	Fetal kidney	48-7-3-G5-PU
			Prostate	
			Hypertrophic prostate	
			Spleen	
			Lung (cells)	
			Umbilical cord	
			Testis	
			Brain	
			Lymph ganglia	
ID136	est-not-ext	8.6	Fetal kidney	78-17-2-E5-PU
			Normal prostate	
ID137	est-not-ext	8.6	Placenta	33-10-4-E2-PU
			Brain	
ID138	est-not-ext	8.5	Umbilical cord	37-11-1-C7-PU
			Normal prostate	

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HELNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID139	est-not-ext	8.5	Fetal kidney Lymphocytes Ovary	26-48-1-H10-PU
ID140	est-not-ext	8.3	Hypertrophic prostate Prostate Cancerous prostate Spleen Normal prostate Brain Lymph ganglia Large intestine	60-13-3-F6-PU
ID141	est-not-ext	8.3	Cancerous prostate Normal prostate	78-22-4-A12-PU
ID142	est-not-ext	8.1	Fetal kidney Ovary Dystrophic muscle Hypertrophic prostate Cancerous prostate Lung Spleen Placenta Fetal brain Normal prostate Colon Brain Substantia nigra	57-28-4-B11-PU
ID143	est-not-ext	8	Cancerous prostate Uterus Lung (cells) Colon Brain Substantia nigra	33-106-3-D8-PU
ID144	est-not-ext	7.9	Normal prostate Colon	23-8-3-F5-PU
ID145	est-not-ext	7.8	Placenta Brain	17-1-3-H5
ID146	est-not-ext	7.6	Lung Normal prostate Brain Substantia nigra	33-37-2-G9-PU
ID147	est-not-ext	7.6	Brain Testis	51-16-4-H4-PU
ID148	est-not-ext	7.6	Hypertrophic prostate Cancerous prostate Fetal brain Muscle Brain Lymph ganglia Large intestine Surrenals	33-32-3-G1-PU
ID149	est-not-ext	7.6	Fetal kidney	47-10-4-F3-PU

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HEIJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID150	est-not-ext	7.4	Hypertrophic prostate	51-1-3-G10-PU
			Cancerous prostate	
			Lung (cells)	
			Umbilical cord	
			Normal prostate	
ID151	est-not-ext	7.4	Brain	33-39-4-B2-PU
			Surrenals	
			Substantia nigra	
			Heart	
			Cancerous prostate	
ID152	est-not-ext	7.4	Testis	47-14-3-A3-PU
			Umbilical cord	
			Brain	
			Lymph ganglia	
			Normal prostate	
ID153	est-not-ext	7.4	Brain	48-53-3-H11-PU
			Substantia nigra	
			Liver	
			Lymph ganglia	
			Cerebellum	
ID154	est-not-ext	7.4	Dystrophic muscle	33-63-1-C3-PU
			Hypertrophic prostate	
			Heart	
			Uterus	
			Umbilical cord	
ID155	est-not-ext	7.3	Brain	53-3-4-F11-PU
			Fetal kidney	
			Ovary	
			Hypertrophic prostate	
			Spleen	
ID156	est-not-ext	7.2	Lung (cells)	48-5-4-E8-PU
			Umbilical cord	
			Normal prostate	
			Brain	
			Substantia nigra	
ID157	est-not-ext	7.1	Fetal kidney	48-54-3-D2-PU
			Fetal brain	
			Uterus	
			Muscle	
			Umbilical cord	
ID157	est-not-ext	7.1	Lung (cells)	48-54-3-D2-PU
			Colon	
			Normal prostate	
			Brain	
			Lymph ganglia	
ID157	est-not-ext	7.1	Fetal liver	48-54-3-D2-PU
			Substantia nigra	
			Surrenals	
			Cancerous prostate	
			Lymph ganglia	
ID157	est-not-ext	7.1	Large intestine	48-54-3-D2-PU

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HEIJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID158	est-not-ext	7.1	Surrenals Prostate Hypertrophic prostate Cancerous prostate	78-18-3-C8-PU
ID159	est-not-ext	7.1	Normal prostate Normal prostate Testis	51-4-2-E10-PU
ID160	est-not-ext	7	Fetal kidney Lymphocytes Umbilical cord	24-11-1-E4-PU
ID161	est-not-ext	7	Cancerous prostate Brain	76-1-2-B8-PU
ID162	est-not-ext	6.7	Ovary Thyroid Cancerous prostate Uterus Muscle Normal prostate Testis	51-11-3-G9-PU
ID163	est-not-ext	6.7	Lymph ganglia Hypertrophic prostate Lung Brain Surrenals	77-16-4-G3-PU
ID164	est-not-ext	6.6	Fetal kidney Hypertrophic prostate	77-38-2-D5-PU
ID165	est-not-ext	6.6	Fetal kidney Cancerous prostate Brain	58-3-3-C8-PU
ID166	est-not-ext	6.5	Brain Testis	51-1-4-C1-PU
ID167	est-not-ext	6.5	Fetal kidney Brain Lymph ganglia	58-9-2-A6-PU
ID168	est-not-ext	6.3	Fetal kidney Cancerous prostate Lung (cells)	30-4-1-E7-PU
ID169	est-not-ext	6.3	Normal prostate Brain	33-51-3-H4-PU
ID170	est-not-ext	6.3	Cancerous prostate Fetal brain	57-27-3-A11-PU
ID171	est-not-ext	6.3	Hypertrophic prostate Fetal brain Normal prostate Brain	57-5-4-G1-PU
ID172	est-not-ext	6.2	Fetal kidney Normal prostate Testis	58-6-1-H4-PU
ID173	est-not-ext	6.2	Fetal kidney Liver Cancerous prostate	37-12-1-D7-PU

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HELJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID174	est-not-ext	6.2	Umbilical cord Cancerous prostate Normal prostate	78-13-1-H1-PU
ID175	est-not-ext	6.2	Large intestine Brain	33-18-3-G10-PU
ID176	est-not-ext	6.2	Substantia nigra Normal prostate	78-39-4-B9-PU
ID177	est-not-ext	6.2	Substantia nigra Brain	33-18-2-B1-PU
ID178	est-not-ext	6.1	Substantia nigra Fetal kidney	37-4-3-D5-PU
ID179	est-not-ext	6.1	Umbilical cord Normal prostate Cerebellum	58-35-3-D12-PU
			Muscle Brain Substantia nigra Fetal kidney Prostate Hypertrophic prostate Cancerous prostate Lung Lung (cells) Umbilical cord Normal prostate Testis Lymph ganglia Large intestine Surrenals	
ID180	est-not-ext	6.1	Fetal liver	51-38-3-D10-PU
ID181	est-not-ext	6.1	Testis Uterus Fetal liver Substantia nigra Ovary Cancerous prostate Fetal brain Normal prostate Lymph ganglia	76-14-3-G2-PU
ID182	est-not-ext	6.1	Cancerous prostate Normal prostate	76-30-1-F7-PU
ID183	est-not-ext	6	Brain	76-43-3-E11-PU
ID184	est-not-ext	6	Cancerous prostate Thyroid Pancreas Fetal kidney Normal prostate	78-41-2-H7-PU
ID185	est-not-ext	5.9	Liver	59-8-1-B7-PU
ID186	est-not-ext	5.8	Lung Brain Lung	78-37-4-E6-PU

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HEIJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID187	est-not-ext	5.8	Normal prostate Kidney Cancerous prostate	59-1-2-E4-PU
ID188	est-not-ext	5.7	Lung Umbilical cord	78-38-4-G2-PU
ID189	est-not-ext	5.7	Normal prostate Lymphocytes Spleen Uterus Substantia nigra Fetal kidney Hypertrophic prostate Cancerous prostate Normal prostate Testis	20-1-3-G5-PU
ID190	est-not-ext	5.7	Brain Fetal kidney	58-37-3-E3-PU
ID191	est-not-ext	5.7	Brain Fetal brain	33-15-1-H3-PU
ID192	est-not-ext	5.6	Lymphocytes Thyroid Spleen Uterus Substantia nigra Hypertrophic prostate Umbilical cord Normal prostate Surrenals	37-1-1-C2-PU
ID193	est-not-ext	5.6	Fetal kidney Umbilical cord Lymph ganglia	48-10-1-A8-PU
ID194	est-not-ext	5.6	Surrenals	62-1-2-D2-PU
ID195	est-not-ext	5.6	Brain Hypertrophic prostate	33-12-4-A7-PU
ID196	est-not-ext	5.6	Brain Normal prostate	78-30-4-H3-PU
ID197	est-not-ext	5.6	Cerebellum Brain Substantia nigra Fetal kidney Hypertrophic prostate Lung Fetal brain Normal prostate Lymph ganglia	47-8-4-C11-PU
ID198	est-not-ext	5.6	Thyroid Brain	84-4-2-C1-PU
ID199	est-not-ext	5.6	Brain Dystrophic muscle Lung (cells) Normal prostate	30-12-4-C2-PU

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HEIJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID200	est-not-ext	5.6	Testis Placenta Lung	1-32-0-D10
ID201	est-not-ext	5.5	Ovary Lung (cells)	30-1-2-E3-PU
ID202	est-not-ext	5.5	Ovary Prostate Lymph ganglia	60-11-1-F1-PU
ID203	est-not-ext	5.5	Spleen Brain Fetal kidney Prostate Hypertrophic prostate Lung (cells) Umbilical cord Testis Lymph ganglia	33-105-2-C3-PU
ID204	est-not-ext	5.5	Cancerous prostate Normal prostate	76-31-4-H1-PU
ID205	est-not-ext	5.5	Fetal kidney Ovary Cancerous prostate Umbilical cord Lung (cells)	30-10-3-B10-PU
ID206	est-not-ext	5.4	Muscle Fetal kidney Cancerous prostate Lung Lymph ganglia	27-3-2-E11-PU
ID207	est-not-ext	5.3	Placenta Muscle Brain Substantia nigra Cancerous prostate Umbilical cord	31-9-2-F9-PU
ID208	est-not-ext	5.3	Brain Substantia nigra Fetal kidney	47-40-3-D2-PU
ID209	est-not-ext	5.3	Brain Substantia nigra Lung	33-77-1-F10-PU
ID210	est-not-ext	5.2	Cerebellum Ovary Umbilical cord Testis	51-19-3-D6-PU
ID211	est-not-ext	5.2	Brain Hypertrophic prostate Colon Testis	51-6-2-F10-PU
ID212	est-not-ext	5.2	Brain Fetal kidney	33-72-4-C5-PU

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HEIJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
			Fetal brain	
			Umbilical cord	
ID213	est-not-ext	5	Normal prostate	33-18-3-E6-PU
			Brain	
ID214	est-not-ext	5	Normal prostate	33-5-2-E1-PU
			Brain	
			Substantia nigra	
			Fetal kidney	
			Umbilical cord	
ID215	est-not-ext	5	Lymph ganglia	76-22-3-E4-PU
			Liver	
			Uterus	
			Muscle	
			Heart	
ID216	est-not-ext	5	Cancerous prostate	51-15-2-H5-PU
			Fetal kidney	
ID217	est-not-ext	4.9	Testis	78-33-3-A9-PU
			Colon	
ID218	est-not-ext	4.9	Normal prostate	58-42-2-H11-PU
			Brain	
			Substantia nigra	
			Fetal kidney	
			Dystrophic muscle	
			Cancerous prostate	
			Lung	
ID219	est-not-ext	4.9	Lymph ganglia	33-111-3-F7-PU
			Brain	
ID220	est-not-ext	4.9	Substantia nigra	76-44-3-C5-PU
			Substantia nigra	
			Fetal kidney	
			Hypertrophic prostate	
ID221	est-not-ext	4.9	Cancerous prostate	78-40-4-B10-PU
			Substantia nigra	
			Normal prostate	
			Testis	
ID222	est-not-ext	4.9	Surrenals	78-6-3-F5-PU
			Fetal kidney	
ID223	est-not-ext	4.9	Normal prostate	58-48-4-E2-PU
			Thyroid	
			Brain	
ID224	est-not-ext	4.8	Fetal kidney	77-38-1-F10-PU
			Placenta	
			Hypertrophic prostate	
ID225	est-not-ext	4.8	Normal prostate	30-7-4-D6-PU
			Lung (cells)	
ID226	est-not-ext	4.8	Normal prostate	48-4-2-H3-PU
			Cancerous prostate	
ID227	est-not-ext	4.8	Lymph ganglia	33-77-4-E8-PU
			Brain	
			Dystrophic muscle	
			Normal prostate	

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HEIJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID228	est-not-ext	4.8	Brain	33-111-2-B4-PU
ID229	est-not-ext	4.7	Substantia nigra Normal prostate Surrenals	62-8-1-A5-PU
ID230	est-not-ext	4.7	Brain	33-6-1-G11-PU
ID231	est-not-ext	4.7	Fetal kidney Fetal liver Substantia nigra Fetal kidney Heart Cancerous prostate Umbilical cord Normal prostate	58-13-1-H2-PU
ID232	est-not-ext	4.7	Liver Brain Substantia nigra Fetal kidney Lung (cells) Testis Large intestine	58-40-2-H6-PU
ID233	est-not-ext	4.7	Brain	33-50-3-C3-PU
ID234	est-not-ext	4.7	Fetal brain Thyroid Spleen Placenta Muscle Brain Substantia nigra Fetal kidney Ovary Heart Cancerous prostate Lung Fetal brain Umbilical cord Normal prostate Colon Testis Lymph ganglia Surrenals	62-10-4-C5-PU
ID235	est-not-ext	4.6	Prostate Lung (cells)	60-16-2-F2-PU
ID236	est-not-ext	4.6	Muscle Brain Substantia nigra Fetal brain Testis	33-87-2-D2-PU
ID237	est-not-ext	4.6	Liver	33-80-3-B8-PU
ID238	est-not-ext	4.5	Brain Liver Cancerous prostate	22-12-3-D4-PU

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HEIJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID239	est-not-ext	4.5	Normal prostate Lymphocytes Spleen Uterus Placenta Muscle Brain Substantia nigra Fetal kidney Ovary Prostate Dystrophic muscle Hypertrophic prostate Heart Cancerous prostate Lung Fetal brain Lung (cells) Umbilical cord Normal prostate Colon Testis Lymph ganglia Surrenals	48-51-4-C11-PU
ID240	est-not-ext	4.5	Cerebellum Substantia nigra	47-15-1-H8-PU
ID241	est-not-ext	4.4	Normal prostate Hypertrophic prostate Lung (cells)	30-12-3-G5-PU
ID242	est-not-ext	4.4	Brain Fetal kidney Cancerous prostate Umbilical cord Normal prostate	58-4-4-D4-PU
ID243	est-not-ext	4.4	Spleen	53-3-2-D4-PU
ID244	est-not-ext	4.4	Pancreas Fetal kidney	58-54-2-H8-PU
ID245	est-not-ext	4.4	Thyroid Kidney Muscle Brain Ovary Cancerous prostate Umbilical cord Normal prostate	27-17-2-C12-PU
ID246	est-not-ext	4.4	Liver Placenta Heart Normal prostate Lymph ganglia	48-5-3-A1-PU
ID247	est-not-ext	4.4	Placenta	33-21-3-D12-PU

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HELINE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID248	est-not-ext	4.4	Brain Substantia nigra	47-2-3-B3-PU
ID249	est-not-ext	4.3	Fetal kidney Umbilical cord Muscle	58-15-2-D7-PU
ID250	est-not-ext	4.3	Fetal kidney Cancerous prostate Lung (cells)	58-41-1-G7-PU
ID251	est-not-ext	4.2	Substantia nigra Fetal kidney Fetal brain Brain	77-5-3-F3-PU
ID252	est-not-ext	4.2	Fetal kidney Hypertrophic prostate Normal prostate Brain	33-106-2-B3-PU
ID253	est-not-ext	4.2	Fetal kidney	58-3-3-B2-PU
ID254	est-not-ext	4.2	Normal prostate	48-46-2-G12-PU
ID255	est-not-ext	4.1	Lymph ganglia Brain Substantia nigra Fetal kidney Hypertrophic prostate Lung (cells) Testis	58-44-2-B3-PU
ID256	est-not-ext	4.1	Cerebellum	47-18-4-E3-PU
ID257	est-not-ext	4.1	Substantia nigra Muscle	78-21-3-F8-PU
ID258	est-not-ext	4.1	Substantia nigra Normal prostate Brain	33-49-1-H4-PU
ID259	est-not-ext	4.1	Surrenals Brain Fetal kidney Fetal brain Normal prostate Colon	23-11-1-E11-PU
ID260	est-not-ext	4	Cerebellum Brain Heart Fetal brain Normal prostate	33-5-2-H4-PU
ID261	est-not-ext	4	Brain	78-12-4-D9-PU
ID262	est-not-ext	4	Normal prostate Spleen Brain	33-103-1-D10-PU
ID263	est-not-ext	4	Hypertrophic prostate Normal prostate Placenta Brain	33-100-4-B7-PU

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HEIJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID264	est-not-ext	3.9	Substantia nigra Hypertrophic prostate Dystrophic muscle Umbilical cord	29-11-2-D6-PU
ID265	est-not-ext	3.9	Normal prostate	78-27-3-D1-PU
ID266	est-not-ext	3.9	Brain Hypertrophic prostate Cancerous prostate	76-30-1-H7-PU
ID267	est-not-ext	3.9	Uterus	74-10-3-C9-PU
ID268	est-not-ext	3.9	Substantia nigra Hypertrophic prostate Cancerous prostate	76-19-1-A9-PU
ID269	est-not-ext	3.9	Liver Muscle Brain Cancerous prostate Normal prostate	76-44-4-A6-PU
ID270	est-not-ext	3.8	Uterus Brain	74-2-1-H4-PU
ID271	est-not-ext	3.8	Substantia nigra Muscle Lung (cells)	27-21-1-H3-PU
ID272	est-not-ext	3.8	Placenta Brain	33-13-3-E8-PU
ID273	est-not-ext	3.8	Thyroid Brain Heart Cancerous prostate Fetal brain Lung (cells) Normal prostate Testis Lymph ganglia	84-3-1-G10-PU
ID274	est-not-ext	3.7	Uterus Brain Fetal kidney Cancerous prostate Dystrophic muscle Cancerous prostate	33-8-1-A3-PU
ID275	est-not-ext	3.7	Thyroid Placenta	76-43-4-H1-PU
ID276	est-not-ext	3.7	Brain	84-5-4-H7-PU
ID277	est-not-ext	3.7	Lung (cells) Umbilical cord Testis Lymph ganglia	37-4-1-B2-PU
ID278	est-not-ext	3.7	Kidney Placenta Uterus Hypertrophic prostate Normal prostate	74-11-4-A9-PU

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HEIJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID279	est-not-ext	3.7	Lymph ganglia Surrenals Substantia nigra Hypertrophic prostate	77-2-2-B9-PU
ID280	est-not-ext	3.7	Cancerous prostate Fetal kidney Cancerous prostate	58-8-1-F2-PU
ID281	est-not-ext	3.7	Lymph ganglia Uterus Prostate Normal prostate	74-7-2-F2-PU
ID282	est-not-ext	3.6	Lymph ganglia Fetal kidney Umbilical cord Testis Large intestine	37-2-1-H11-PU
ID283	est-not-ext	3.5	Lymphocytes Brain Fetal kidney Normal prostate	58-6-1-F3-PU
ID284	est-not-ext	3.5	Muscle Brain Hypertrophic prostate	33-54-3-G1-PU
ID285	est-not-ext	3.5	Fetal liver Substantia nigra	47-39-2-H6-PU
ID286	est-not-ext	3.5	Brain Cancerous prostate Surrenals	76-17-1-F5-PU
ID287	est-not-ext	3.5	Placenta Muscle Heart Cancerous prostate Lung (cells) Umbilical cord Colon	27-7-3-D1-PU
ID288	est-not-ext	3.5	Liver Uterus Muscle Brain Ovary Dystrophic muscle Cancerous prostate Normal prostate Colon Large intestine	74-5-1-E4-PU
ID289	est-not-ext	3.5	Brain Cancerous prostate Fetal brain Umbilical cord Surrenals	57-20-1-F6-PU

<u>SEQ. ID NO.</u>	<u>CATEGORY</u>	<u>VON HEIJNE SCORE</u>	<u>TISSUE SOURCE</u>	<u>INTERNAL DESIGNATION</u>
ID290	ext-vrt-not-genomic	7.4	Spleen Hypertrophic prostate Lymph ganglia	48-25-3-A3-PU
ID291	ext-vrt-not-genomic	7	Brain Pancreas Hypertrophic prostate Normal prostate	46-1-3-F4-PU

TABLE III

SEQ. ID NO.	SIGNAL PEPTIDE
ID38	MSSWSRQRPKSPGGIQPHVSRTLFLLLLLAASAWG
ID39	MRVRIGLTLLLXAVLLSLASA
ID40	MFSHLPFDCVLLLLLLLLLTRS
ID41	MGPVRLGILLFLFLAVDEAWA
ID42	MKSLSLLAVALGLATA
ID43	MLLLTLXLGGPTWA
ID44	MKIGILLSLLNSVISQTLMSCNWKQMRMKTILILIXIWIWCLG
ID45	MKASSGRCGLVRWLQVLLPFLSLFPGALP
ID46	MTVDCVSSHLKKTGDGAKTFHFLCHLLRGLHA
ID47	MAKALLFPSGRSVRLYGAVNKERQXESVLNRACPPKANSKERRGRAVLGAELTQWSSPT TAGSCCSSLTLCARSSSXVIAPSPLVPFTSGLTSLSWLLXASCS
ID48	MAASEAAVVSSPSLKTDTSPVLETAGTVAAMAATPSARAAAAVAAAAARTGSEARVS KAALATKLLSLSGVFA
ID49	MKVGVLWLISFFTFIDG
ID50	MEFGLSWIFLAAILKGVQC
ID51	MAEPGHSHLSARVRGRTERIPRLWRLLLWAGTAFQ
ID52	MTADPRKGRMGLQACLLGLFALIS
ID53	MLVDGPSERPALCFLLAVAMSFF
ID54	MAAPLVVLVAVTVRA
ID55	MTAARRQRELSILPKVTLEAMNTTVMQGFNRSECRDTRIVQLVFPALYTVVFLTGIL LNTLALWVFVHIPSSSTFIYLNKTLVADLXMTLMLPFKILS
ID56	MSSVLAASHPLVLSSNAGTPGISEKDNRPAGSSIGVLTSLHLSG
ID57	MGLAMEHGGSYARAGSSRGWYYLRYFFLFVSLIQFLILGLVLFMVG
ID58	MVEASLSVRHPEYNRPLLANDLMLIKLDESSESITIRSISIASQCPTAGNSCLVSGWGL LANG
ID59	MGGKQRDEDEAYGKPVKYDPSFRGPIKNRSCTDVICCVLFLFILG
ID60	MQKASVLLFLAWVCFLFY
ID61	MSPVLHFYVRPSGHEGAASGHTRRKLQGKLPELQGVETELCYNVNWTAELPSAEETKKL MWLFGCPYCWMMMLGSXGSFL
ID62	MDVTPRESLSILVVAGSGGHTTEILRLLGSLSNAYS
ID63	MMGVAKLTLLRVNLPHNSIG
ID64	MDVTPRESLSILVVAGSGGHTTEILRLLGSLSNAYS
ID65	MVLLTMIARVADG
ID66	MVPVENTEGPSLLNQKGTAVETEGXGSRHPPWARGCGMFTFLSSVXA
ID67	METFLEPNKLLFPVGRSWSCFA
ID68	MGFLWGLALPLFFFC
ID69	MQSTSNHLWLLSDILGQGATA
ID70	MVEICAGSVLPYSNC
ID71	MVAPVLETSHVFCCPNRVRGVNLWXSGPRGLLAFGTSCSVVXY
ID72	MDSLRKMLISVAMLGAXAGVGYALLVIVTPGERRKQEMLKEMPLQDPRSREEAART QQLLLATLQEAATT
ID73	MRQTLPCIYFWGGLLPFGMLCASSTT
ID74	MADDLEQQSQGWLSSWLPTWRPTSMSQLKNVEARILQCLQNKFLARYVSLPNQNKI WTVTVSPEQNDRTPLVMVHGFGGVGVLWLNMDSLXARRTLHTXGLLGFRXQG
ID75	MKVTGITILFWPLSMILLSDKIQS
ID76	MAAGRAQVPSSEQAWLEDAQVFIQKTLCPAVKEPNVQLTPLVIDCVKTVWLSQGRN QGSTPLSYSFVSVDLKTQHRLPCCSHLSWSSSAYQAWA
ID77	MSTCCWCTPGGAST
ID78	MPFAEDKTYKYICRNFNFSCXVDVVEILPYLPCLTA

<u>SEQ. ID</u> <u>NO.</u>	<u>SIGNAL PEPTIDE</u>
ID79	MAESEDRLRIVLVGKTGSGKSATANTILGEEIFDSRIAQAQAVTKNCQKASREWQGRDLL VVDTPGLFDTKESLXTTCKEIXRCIISSCPGPHAIVLVLLGRYTEE
ID80	MAQKPLRLLACGDVEGKFDILFNRVQAIQKXSGNFDLLXCVGNFFGSTQ
ID81	MESRKDITNQEELWKMKPRRNLEEDDYLHKDTGETSMLKRPVLLHLHQTAAH
ID82	MESRKDITNQEEXWKMKPRRNLEEDDYLHKDTGETSMLKRPVLLHLHQTAAH
ID83	MAATCEISNIFSNYFSAMYSSDSTLASVPPAATFG
ID84	MRDCPGVEXILDCSXRQKTEGCLQAGKECVDSPVEGGQSEAPPSLVSFVSSGTEQ
ID85	MERQSRVMSEKDEYQFQHQGAVELLVFNFLILILT
ID86	MKMASSLAFLLLNFHVSLLLQLLTPCSA
ID87	MVFLPLKWSLATMSFLLSLLALLTVSTPSWC
ID88	MESAAALHFSRPASLLLLLXCXVHWS
ID89	MEKIPVSAFLLLVALSYTLA
ID90	MGPWGEPELLVWRPEAVASEPPVPVGLVKGALVLLVLTLLCSL
ID91	MAPLLLQLAVLGAALA
ID92	MAMEGYWRFLXLLGSALLVGFLSVIFA
ID93	MAQSLALSLLLVLAFG
ID94	MEAMWLLCVALAVLA
ID95	MAPITTSREEFDEIPTVVGIFSAFGLVFTVSLFAWICC
ID96	MEGPRGWLVLCLAISLA
ID97	MTAWEAMAPHVNPTLKDALKSPQQXXXTSPAPCXSNHHNKKHLILAFCAVLLTLLLIAF IFL
ID98	MLCSLLLCECLLLXAGYA
ID99	MGHAMGLVXSLPVHCLTFA
ID100	MARCFSLVLLTTSIWT
ID101	MLLTRKQTCQLGILLSIHRQHSKDLQDIVATLGPRSAATHPHQPAIQVLAQLAFLSQISQ
ID102	MWAFSELPMPLLINLIVSLLGFVATVTL
ID103	MFKVIQRSVGPASLSLLTFKVYA
ID104	MAKSLLKTASLSGRTKLLHQTGLSLYSTSHGFYEEEVKKTLLQFPGGSIDLQKEDNGIGI LTLNPNPSRMNAFSGVMMLQLLEKVIENWTEGKGLIVRGAKNTFSSGSDLNAVKSGLGQ RLPLISVALVQGWALG
ID105	MTSFSTAQCSTSDSACRISPGQINXVRPKLPLLKILHAAGAQQ
ID106	MDTAEDICRVCRSEGTPEKPLYHPCVCTGSIKXVHQECLVQWLKHSRKEYCELCKHRFA FTPIYSPDMP SRLPIQDIFAGLVTSIGTAIRYWFHYTLVAFWLGVPPLTAC
ID107	MLIMLGIFFNVHS
ID108	MGGLWRPGWRCVPFCGWRWIHPGSPTRAERVEPFLRPEWSGTGAERGLRWLGTWKR CSLRARHPALQPPRRPKSSNPFTAXEEERRRXNKTTLTYYAAVAVGMLXASYA
ID109	MAAQCVTKVALNVSCANLLDKDIGSKSDPLCVLFLNTSG
ID110	MTGSNEFKLNQPPEDGISSVKFSPNTSQFLLVSSWDTSVRLYDVPANSMRLKYQHTGAVL DCAFYDPTHA
ID111	MGKHLWYPGQASAHLCWCGSHCCST
ID112	MLAVSLTVXLLGA
ID113	MSSTLAKIAEIEAEMARTQKNKATAHHLGLLKARLAKLRRELITPKGGGGGGPGEGFDWP RQVMLELDLLVFHLWG
ID114	MAAAVPKMRGPAQAKLLPGSAIQALVGLARPLVLALLVSAALS
ID115	MTPQSLLQTTLFLLSLLFLVQGAHG
ID116	MMVVGTTGSLALSSLLSLLFAGMQIYSRQLASTEWTIQQGGLLGSGLFVFSLTAFNNLE NLVFGKGFQAKIFPEILLCLLLALFASG
ID117	MDWTWRVFCLLAVAPGAHS
ID118	MRIANRTRFSSPFLARGAGWTHGRGMMVVGTTGSLALXSLLSLLFAGMQMYSRQLASTE WLTIQGGLLGSGLFVFSLTAFNNLENLVFGKGFQAKIFPEILLCLLLALFASG
ID119	MTSVSTQLSLVMSLLLVLVVEA

SEQ. ID NO.	SIGNAL PEPTIDE
ID120	MTPLLTLVVLMLPLAQA
ID121	MALLALSLLVLTSP
ID122	MGGLEPCSRLLPLLLAVSG
ID123	MEVPPAPRSFLCRALCLFPRVFA
ID124	MDLRQFLMCLSLCTAFALS
ID125	MAGGVRPLRGLRALCRVLLFLSQFCILSGG
ID126	MAAAAWLQVLPVILLLLGAHP
ID127	MRTLFNLLWLALACSPVHT
ID128	MDVLFVAIFAVPLLG
ID129	MAAAAWLQVLPVILLLLGAHP
ID130	MRTLFNLLXLALACSPVHT
ID131	MGSKVADLLYWKDTRTSGVVFTGLMVSLCLLHFSIVVA
ID132	MAARWRFWCVSVTMVVALLIVCDVPSASA
ID133	MEGESTSAVLSGFVLGALA
ID134	MFAPAVMRAFRKNKTLGYGVPMLLIVGGSFG
ID135	MAAAWXSGPSAPEAVTARLVGVLFVSVTTGPWGAVATSAGGEESLKCEDLVGQ YICKDPKINDATQEPVNCTNYTAHVSCFPAPNITCKDSSGNETHFTGNEVGFFKPISCRNV NGYSYKVAVALSLFLGWLGA
ID136	MRTLFNLLWLALACSPVHT
ID137	MDGQKKNWKDKVVDLLYWRDIKKTGVVFGASLFLLSLTVFS
ID138	MVAPGLVLGLVPLILWA
ID139	MSPSGRLCLLTIVGLILPTRG
ID140	MRIANRTRFSLPFLARGAGWTHGRGMMVVGTTSLALSSLLSLLFA
ID141	MVLGGCPVSYLLCGQAALLGNLLHCVSRSHS
ID142	MGSVLGLCSMASWIPCLCGSAPCLLCRCCPSGNNSTVTRLIYALFLLVGVCA
ID143	MVLLHVLFEHAVGYALLALKEVEEISLLQPQVEESVLNLGKFHSIVRLVAFCPFASS
ID144	MSGGRAPAVLLGGVASLLSFVWMPALLPVASRLLLPVLLTMSAG
ID145	MVAPVWYLVAALLVGFIPLTRSRG
ID146	MAVLAPLIALVYSVPRLSRWLAQPYLLSALLSAAFLLVKLPPLCHG
ID147	MVGEAGRDLRRRRXXAVTAXKMAVLAPLIALVYSVPRLSRWLAQPYLLSALLSAAFLLV RKLPLCHG
ID148	MEALGKLKQFDAYPKTLEDVRVKTCCGATVTIVSGLLMLLFLSELQY
ID149	MAVLAPLIALVYSVPRLSRWLAQPYLLSALLSAAFLLVKLPPLCHG
ID150	MRCLTTPMLLRALAQAARA
ID151	MRCLTTPMLLRALAQAARA
ID152	MDFITSTAILPLFGCLGVFG
ID153	MHPAVFLSLPDLRCSLLLVTVWVFTPTT
ID154	MASLGHILVFCVGLLTMAKA
ID155	MSGSSLPSALALSLLVSGSLLP
ID156	MAVHDLIFWRDVKKTGTVFGTTLIMLLSLAAFSVIS
ID157	MXGSVECTXGWGHCAPSPLLLWTLFLFAAPFG
ID158	MQCFSEIKTMMILFNLLIFLCGAALLAVG
ID159	MRGSVECTWGXGHCAPSPLLLWTLFLFAAPFG
ID160	MALRLLKLAATSASA
ID161	MPSAFVSFFPVSPAFLVTQTDWTEPWLMLATFHALCVLLTCLSSRSYRLQIGHFLCLV ILVYC
ID162	MALPHQEPKPGDLIEIFRLGYEHWALYXDGVIHLAPPSEYPGAGSSSVFSLNSAEV KRERLEDVGGCCYRVNNSLDHEYQPRPVEVISSAKEMVGQKMKYSIVSRNCEHFVTQL RYGKSRCKQVEKAKVEVGVATAGLVVAGCSFA
ID163	MAASTSMVPVAVTAAVAVLSINSDFSLREIKKQLLLIAGLTRERGLLHSSKWSAELAF SLPALPLAEL

<u>SEQ. ID</u> <u>NO.</u>	<u>SIGNAL PEPTIDE</u>
ID164	MEEGGNLGGLIKMVHLLVLSGAWG
ID165	MAGPAAAFRRLGALSGAAALGFASYGAHGAXFPDAYGKELFDKANKHHFLHSLALL GVPHCRKPLWAGLLASGTTLFCTS
ID166	MGHRFLRGLLTLLPPPLYT
ID167	MELLQVTILFLLPSICSSNS
ID168	MASSNTVLMRLVASAYSIA
ID169	MRSSCVLLTALVALA
ID170	MGIQTSPVLLASLGVGLVTLLGLAVG
ID171	MTLQWAAVATFLYAEIGLILFCLPFIPPQRWQKIFSFNVWGKIATFWNKAFLTIIILLI VLFLDAVRE
ID172	MPSEGRCWETLALRSSDKGRLCYYRDWLLRREVSGGPGGRRPFRPLATETFSLAVGTFC SREPVSNNLHLFLDFCVIPLSWG
ID173	MTKLAQWLWGLAILGSTWVALTTG
ID174	MLLAWVQAFLVSNMLLAEAYG
ID175	MAMHFIFSDTAVLLFHFWSVHSPAGMALSVLVLLLAVALYE
ID176	MKQVHQCIERCHVPLAQALVTSELEKFQDRLARCTMHCNDKAKDSIDAGXKELQ VKQQLXVVXXSVLXTTCXS
ID177	MQMSYAIRCAFYQLLAALMLVAMLQL
ID178	MMTQTCIILLIHTMQVCTT
ID179	MXXHLQTRPLFLTCLFWPLAAL
ID180	MAANYSSTXTRREHVKVKTSSQPGFLERLSETSGGMFVGLMAFLLSFYLIFT
ID181	MRGAHLTALEMLTAFASHIRA
ID182	MVHKPMMTQTCIILLIHTMQVCTT
ID183	MAGIKALISLSFGGAIGLMFLMLGCALP
ID184	MSLMPKMHLFPLTLVRSFWS
ID185	MMKRAAAAAVGGALAVGAVPVVLSAMGFTGAGIAASSIAAKMMSAAAIANGGGVSA GSLVATLQSVGAAGLSTSSNILLASVGSVLG
ID186	MVTIILLSCXFWA
ID187	MXKRAAAAAVGGALAVGAVPVVLSAMGFTGAGIAASSIAAKMMSAAAIANGGGVSA GSLVATLQSVGAAGLSTSSNILLASVGSVSG
ID188	MSQDGGXGELKHMVMSFRVSELQVLLGXXGRNKSGRKHELLAKALHLLKSSCAPSVQ MKIKELYRRRFPRKTLGPSDLSSLPPGTSP
ID189	MPXLLPVASRLLLLPRVLLTMSG
ID190	MVFSNNDEGLINKKLPKELLRIFSFLDIVTLCRC
ID191	MVFSNNDEGLINKKLPKELLRIFSFLDIVTLCRC
ID192	MASYFDEHDCEPSDPEQETRTNMLELARSFNRMDFEDLGLVVDWDHHLPPPAAKTVVE NLPRTVIRGSQAEKCPVCLLEFEEEEETAIEMPCHHLFHSSCILPWLSKTNS
ID193	MPLILSLQVCRPATL
ID194	MLGITSCSDQAKEGEGLEGSSSTGSSSGNHGGSGGGNGHKPGCEKPGNEARGSGNLGFRT LRLLGCLTLTSL
ID195	MARKALKLASWTSMALA
ID196	MAAAALPAWLSLQSRA
ID197	MVKIAFNTPTAVQKEEARQDVEALLSRTVRTQILTGKELRVATQEKEGSSGRCMTLXXL SFILA
ID198	MIGSGLAGSGGAGGPSSTVTWCALXSNHVAATQASLLSFVWMPALLP
ID199	MSGAQLXGFLFXVIVLTS
ID200	MSFFQLLMKRKELIPLVVFMTVAASGASS
ID201	MELAHSLLLNEEALA
ID202	MTSALTQGLERIPDQLGYLVLSEGAVLA
ID203	MAAAWPSGPXAPEAVTARLVGLWVFSVTTG
ID204	MVLLTMIARVADG

SEQ. ID NO.	SIGNAL PEPTIDE
ID205	MVLLTMIARVADG
ID206	MTSQVPVNETIIVLPSNVINFSQAEKPEPTNQGDLSKKHLHAEIKVIGTIQILCGMMVL SLGDXLASA
ID207	MASVVLALRTRTAVTSLLSPTPATA
ID208	MASVVLALRTRTAVTSLLSPTPATA
ID209	MMPSTRNLATGIPSSKVYSRLSSTDGIDLQFKKTPPKIPYKAIALATVFLIGA
ID210	MPLILSLQVCRPATL
ID211	MPLILSLQVCRPATL
ID212	MASSVGNVADSTEPTKRMLSFGLAELAHREYQAGDFEAAERHCMQLWRQEPDNTG VLLLLSSIHQC
ID213	MFGSAPQRPVAMTTAQRDSLLWKLAGLLREXGDVVLSGCSTLSLLTPTLQQLNHVFELHL GPWGPQGTGFVALPSHPADSPVILQQLFDVLQ
ID214	MSFIFEWIYNGFSSVLQFLGLYKKSGLK VFLGLDNAGKTTLLHMLKDDRLGQHVP TLHPT SEELTIAGMTLQLLILVGTSKHVAFG
ID215	MDKPCGCPPGVCDHGTGDRDPWYSTVGLLPPVRA
ID216	MAAALKCLLTLGRWCPGLGVAPQARALAALVPGVTQ
ID217	MVARVWSLMRFLIKGSVAGGAVYLVDQELLGPSDKSQAALQKAGEVVPAMXQFS QYVCQQTGLQIPQLPAPPKIYFPIRDSWXAGIMTVMSALSVAPSKA
ID218	MVNELQNLXSLQGSQA
ID219	MLYMSLKYIRAFFFSIQPFLPCSS
ID220	MNLERSVNEEKLNLCKRYLGGFAFLPFLWL VNIFWFFREAFV PAYTEQSQIKGYVWRS AVGFLFWVIVLTSWITIFQ
ID221	MAGELQGTQAPSLRGXGLTSQDSGVNPNNSXRGREAMASGSNWLSGVNVVLVMAYG SLVFVLLFIFVKRQ
ID222	MTGFLLPPASRGTRRSCSRSRKQTRRRRNPSFVASCP TLLPFACVPGASPTTLA
ID223	MEEXSXPLVEFVKVLCTNQVLITARA
ID224	MVRRLXXVVAFAVAPGES
ID225	MAVPGVGLLTRLNLCARRRTRVQRPIVRLLSCPGTVA
ID226	MMAAVPPGLEPWNRVRIPKAGNRSVTVQNPGAALDL CIAAVIKECHLVLSLKSQTLDA
ID227	MASLDRVKVLVLGDSGVGKSSLVHLLCQNQVLG
ID228	MVFPAKRFLVPSMEGVRWAFSCGTWLP SRA
ID229	MASKIGSRRWMLQLIMQLGSVLLTRC
ID230	MLSKGLKRKREEEEEKEPLAVDSWWLDPGHA
ID231	MDYSLAAALTLHGHWG
ID232	MSYITSQEMKCILHWFANWSGPQRRERFLEDLVAKAVPEKLQPXLD SLEQLSVSGADDHLL SLXASYIFGISG
ID233	MPLLCQIEMEYLLLKWQMTMLQSM LCDLVSYPLLPLQQTKEANLDFPKIKVSSVTITPTR WFXLIVYLVVVSFIAS
ID234	MWFEILPGLSVMGVCLLIPGLA
ID235	MEFKLEAHRIVSISLGIYNSRVQRGGIKLHKNLLVSLVLRXPAKS
ID236	MAVLSKEYGFVLLTGAASFIMVAHLAINVSKARKKYKVEYPIMYSTDPENGHIFNCIQRA HQNTLEVYPXFLFFLA VGGVYHPRIASGLGLXDCWT
ID237	MDGHWSAAFSALTVTAMSSWARRRSSRRIPSLPGSPVCWA
ID238	MAQRLLLRFLASVIS
ID239	MASLKPAFVNYFFLLLEVSHELLI
ID240	MNLERSVNEEKLNLCKRYLGGFAFLPFLWL VNIFWFFREAFV PAYTEQSQIKGYVWRS AVGFLFWVIVLTSWITI
ID241	MAQLGAVVAVASSFFCASLFS
ID242	MSLRNLWRDYKVLVFMVPLVGLIHL
ID243	MGWDGCKCLGVFCLLISIPTSA

<u>SEQ. ID</u> <u>NO.</u>	<u>SIGNAL PEPTIDE</u>
ID244	MAASQAVEEMRTAWFWGSLGFAMSILLTFPVTIPVMMMPGTRXGFEXRXFRVDVVH
ID245	MDENSLEFDMVGIDAAIANAFRRILLAEVPTMAVEKVLVYNNTSIVQDEILAHRLGLIPIHA MAASKVKQDMPPPGGYGPIDYKRNLPRLSGYSMLAIGIGTLIYGHWSIMKWNERRRL QIEDFEARIALPLLQA
ID246	MSGFLEGLRCSECIDWGEKRNITIASIAAGVLFFTGWWIIDA
ID247	MMTQEPGIYTWPKEKTRICSACSSVPLPWTVLVFLTFLSIPSFV
ID248	MFLTALLWRGRIPG
ID249	MNQENPPYPGPGFTAPYPPYPQPMGPGXMGGPYPPQGYPYQGYQYQGWQGGPQEPPK TTVYVVEDQRRDELGPSTCLTACWTALCCC
ID250	MASLEVSRSPRRSRRELEVRSPRQNKHSVLLPTYNEREELPLIVWLLVKSFSSES
ID251	MCPTCLCAPSXXWG
ID252	MAAATGAVAASAASGQAEQ
ID253	MAAMSLXRVSVTAVAA
ID254	MAGPLQGGGARALDLLRGLPRVSLA
ID255	MATATEQWVLVEMVQALYEAPAYHLILEGILILWIRLLFS
ID256	MEDPNPEENMXQQDSPKERSQSPGGNICHLGAPKCTRCLITFADSKXXERHMKREHPAD FVAQKLQGVLFICFTCARS
ID257	MNVIDHVRDMAAAGLHSNVRLSSLLTMSNN
ID258	MQNVINTVKGKALEVAEYLTPLVKESFKETGVITPEEFVAAGDHLVHCPTWQWATG
ID259	MATLTFSLRKPLQRSIRPSHLPLCCFDWRLSHYYRLPPAVRLHQQRGGRPGRSSADHWH SGVPTRLPPAHRLLCIQRLPWLLCRG
ID260	MEKPLFPLVPLHWFGGYTALVVS GGIVGYVKTGSVPSLAAGLLFGSXA
ID261	MASTVVAVGLTIAAAGFA
ID262	MVIRVYIASSSGSTAIKKKQQDVLGFLEANKIGFEEKDIAANEENRKWMRENVPENSRPA VQGPFAFRYKAFSFSRLLSQCPR
ID263	MSSRGHSTLPRTLMAPRMISEGDIGGIAQITSSLFLGRGSA
ID264	MAAPGPALCLFDVDGTLT
ID265	MPLGARILFHGVFYAGGFA
ID266	MLLSIGMLMLSAT
ID267	MSLTSSSSVRVEWIAAVTIAAGTAA
ID268	MSGSGSKENSHNKARTSPYPGSKVERSQVPNEKVGWLVEWQDYKPYEYTA VSVLA GPRWA
ID269	MAISLRSSGISVKCLSKLWMRWTVTSTTRA
ID270	MSEVRLPPLRALDDFVLGSARLGSGS
ID271	MKLVSATAWLEECWW
ID272	MKAISVSLRLTKLLWFFSIVLYVPLLA VCCLHS
ID273	MGSLSGRLAAGSCFRLCERDVXSRLRTRSSDLKRINGFCTKPQESPGAPSRTYNRVPL HKPTDWQKKLIWSGRFKKEXXIPETVSLEMLXXAKNKMVRKISYLMIALTVVGCIFM
ID274	METLYRVPFLVLECPNLKLLKPPWLHMP SAMTVYALVVVS YFLITGGIYDVIVEPPSVG SMTDEHGHQRPVAFLAYRVNGQYIMEGLASSFLTMGGLG
ID275	MLVLRSGLTKALA
ID276	MAAPLSVEVEFGGAXSCLTVLRNIESLAWTGGTLG
ID277	MTHLIEYDRHRKSRLSPLQHL YLLPADHSRNAAERFPGAWFQPPTVDSEASAFVGGLPVI FWSWA
ID278	MAAAALGQIWARKLLSVPWLLC
ID279	MAVESRVTQEEIKKEPEKPIDREKTCPLLLLVFTTNNG
ID280	MRLKYQHTGAVLDCAFYDPHTA
ID281	MALLFARSLRLCRWGAKRLGVASTE AQRGV SFKLXECTAHSSSLALFRDDTG VKYGL VGLEPTKVALNVERFREWA VVLADTA VTSG
ID282	MAAAAAGTXTSQRFFQSFS DALIDEDPQA ALEELTKALEQKPDDAQY YCQRAYCHILLGN YCVAVADA

SEQ. ID
NO.

SIGNAL PEPTIDE

ID283	MAQLKYMENVGYAQEDRERMHRNIVSLAQNLLNFMIGSILDLWQCFLWFIYIGSSLNGTRG
ID284	MSPAFRAMDVEPRAKGSFWSPLSTRSGGTHA
ID285	MADEELEALRRQRLAELQAKHGDPGDAAQQEAKHREAEMRNSILAQVLDQSARA
ID286	MSAAGARGLRATYHRLLDKVELMLPEKLRPLYNHPAGPRTVFFWAPIMKWGLVCAGL ADMARP
ID287	MSNYSVSLVGPAPWGFRLQGGKDFNMPLTISSLKDGGKAAQANVRIGDVVLSIDGINAQQ MTHLEAQNKIKGCTGXLNMTLQRASA
ID288	MANPKLLGLELSEAEIG
ID289	MIIPLEILIIIVLNEVLLFDVNSVYKALLCTLLHFNQNI
ID290	MDIQMANNFPPSATPQGNDCLYAAHSTARIVMPLHYSLVFIIGLVGNLLA
ID291	MLTIVKSPQKSYLFPSSMIGIGSLPSCWA

Minimum signal peptide score	false positive rate	false negative rate	proba(0.1)	proba(0.2)
3.5	0.121	0.036	0.467	0.664
4	0.096	0.06	0.519	0.708
4.5	0.078	0.079	0.565	0.745
5	0.062	0.098	0.615	0.782
5.5	0.05	0.127	0.659	0.813
6	0.04	0.163	0.694	0.836
6.5	0.033	0.202	0.725	0.855
7	0.025	0.248	0.763	0.878
7.5	0.021	0.304	0.78	0.889
8	0.015	0.368	0.816	0.909
8.5	0.012	0.418	0.836	0.92
9	0.009	0.512	0.856	0.93
9.5	0.007	0.581	0.863	0.934
10	0.006	0.679	0.835	0.919

TABLE IV

143/2

Minimum signal peptide score	All ESTs	New ESTs	ESTs matching public EST closer than 40 bp from beginning	ESTs extending known mRNA more than 40 bp	ESTs extending public EST more than 40 bp
3.5	2674	947	599	23	150
4	2278	784	499	23	126
4.5	1943	647	425	22	112
5	1657	523	353	21	96
5.5	1417	419	307	19	80
6	1190	340	238	18	68
6.5	1035	280	186	18	60
7	893	219	161	15	48
7.5	753	173	132	12	36
8	636	133	101	11	29
8.5	543	104	83	8	26
9	456	81	63	6	24
9.5	364	57	48	6	18
10	303	47	35	6	15

TABLE V

Tissue	All ESTs	New ESTs	ESTs matching public EST closer than 40 bp from beginning	ESTs extending known mRNA more than 40 bp	ESTs extending public EST more than 40 bp
Brain	329	131	75	3	24
Cancerous prostate	134	40	37	1	6
Cerebellum	17	9	1	0	6
Colon	21	11	4	0	0
Dystrophic muscle	41	18	8	0	1
Fetal brain	70	37	16	0	1
Fetal kidney	227	116	46	1	19
Fetal liver	13	7	2	0	0
Heart	30	15	7	0	1
Hypertrophic prostate	86	23	22	2	2
Kidney	10	7	3	0	0
Large intestine	21	8	4	0	1
Liver	23	9	6	0	0
Lung	24	12	4	0	1
Lung (cells)	57	38	6	0	4
Lymph ganglia	163	60	23	2	12
Lymphocytes	23	6	4	0	2
Muscle	33	16	6	0	4
Normal prostate	181	61	45	7	11
Ovary	90	57	12	1	2
Pancreas	48	11	6	0	1
Placenta	24	5	1	0	0
Prostate	34	16	4	0	2
Spleen	56	28	10	0	1
Substantia nigra	108	47	27	1	6
Surrenals	15	3	3	1	0
Testis	131	68	25	1	8
Thyroid	17	8	2	0	2
Umbilical cord	55	17	12	1	3
Uterus	28	15	3	0	2
Non tissue-specific	568	48	177	2	28
Total	2677	947	601	23	150

TABLE VI

Description of Transcription Factor Binding Sites present on promoters isolated from SignalTag sequences

Promoter sequence P13H2 (646 bp):

Matrix	Position	Orientation	Score	Length	Sequence
CMYB_01	-502	+	0.983	9	TGTCAGTTG
MYOD_Q6	-501	-	0.981	10	CCCAACTGAC
S8_01	-444	-	0.960	11	AATAGAATTAG
S8_01	-425	+	0.968	11	AACTAAATTAG
DELTAEF1_01	-390	-	0.960	11	GCACACCTCAG
GATA_C	-364	-	0.984	11	AGATAAATCCA
CMYB_01	-349	+	0.958	9	CTTCAGTTG
GATA1_02	-343	+	0.959	14	TTGTAGATAGGACA
GATA_C	-339	+	0.953	11	AGATAGGACAT
TAL1ALPHA47_01	-235	+	0.973	18	CATAACAGATGGTAAG
TAL1BETA47_01	-235	+	0.983	18	CATAACAGATGGTAAG
TAL1BETA1F2_01	-235	+	0.978	18	CATAACAGATGGTAAG
MYOD_Q8	-232	-	0.954	10	ACCATCTGTT
GATA1_04	-217	-	0.953	13	TCAAGATAAAGTA
IK1_01	-126	+	0.983	13	AGTTGGGAATTCC
IK2_01	-126	+	0.985	12	AGTTGGGAATTC
CREL_01	-123	+	0.962	10	TGGGAATTCC
GATA1_02	-96	+	0.950	14	TCAGTGATATGGCA
SRY_02	-41	-	0.951	12	TAAACAAAACA
E2F_02	-33	+	0.957	8	TTTAGCGC
MZF1_01	-5	-	0.975	8	TGAGGGGA

Promoter sequence P15B4 (861bp):

Matrix	Position	Orientation	Score	Length	Sequence
NFY_Q8	-748	-	0.956	11	GGACCAATCAT
MZF1_01	-738	+	0.962	8	CCTGGGGA
CMYB_01	-684	+	0.994	9	TGACCGTTG
VMYB_02	-682	-	0.985	9	TCCAACGGT
STAT_01	-673	+	0.968	9	TTCTGGAA
STAT_01	-673	-	0.951	9	TTCCAGGAA
MZF1_01	-556	-	0.956	8	TTGGGGGA
IK2_01	-451	+	0.965	12	GAATGGGATTTC
MZF1_01	-424	+	0.986	8	AGAGGGGA
SRY_02	-398	-	0.955	12	GAAACAAAACA
MZF1_01	-216	+	0.960	8	GAAGGGGA
MYOD_Q6	-190	+	0.981	10	AGCATCTGCC
DELTAEF1_01	-176	+	0.958	11	TCCCACCTTCC
S8_01	5	-	0.992	11	GAGGCAATTAT
MZF1_01	16	-	0.986	8	AGAGGGGA

Promoter sequence P29B6 (655 bp):

Matrix	Position	Orientation	Score	Length	Sequence
ARNT_01	-311	+	0.964	18	GGACTCACGTGCTGCT
NMYC_01	-309	+	0.965	12	ACTCACGTGCTG
USF_01	-309	+	0.985	12	ACTCACGTGCTG
USF_01	-309	-	0.985	12	CAGCACGTGAGT
NMYC_01	-309	-	0.956	12	CAGCACGTGAGT
MYCMAX_02	-309	-	0.972	12	CAGCACGTGAGT
USF_C	-307	+	0.997	8	TCACGTGC
USF_C	-307	-	0.991	8	GCACGTGA
MZF1_01	-292	-	0.968	8	CATGGGGA
ELK1_02	-105	+	0.963	14	CTCTCCGGAAGCCT
CETS1P54_01	-102	+	0.974	10	TCCGGAAGCC
AP1_Q4	-42	-	0.963	11	AGTGACTGAAC
AP1FJ_Q2	-42	-	0.961	11	AGTGACTGAAC
PADS_C	45	+	1.000	9	TGTGGTCTC

TABLE VII

CLAIMS

1. A purified or isolated nucleic acid comprising the sequence of one of SEQ ID NOs: 38-291 or comprising a sequence complementary thereto.
- 5 2. The nucleic acid of Claim 1, wherein said nucleic acid is recombinant.
3. A purified or isolated nucleic acid comprising at least 10 consecutive bases of the sequence of one of SEQ ID NOs: 38-291 or one of the sequences complementary thereto.
4. A purified or isolated nucleic acid comprising at least 15 consecutive bases of
10 one of the sequences of SEQ ID NOs: 38-291 or one of the sequences complementary thereto.
5. The nucleic acid of Claim 4, wherein said nucleic acid is recombinant.
6. A purified or isolated nucleic acid of at least 15 bases capable of hybridizing under stringent conditions to the sequence of one of SEQ ID NOs: 38-291 or one of the
15 sequences complementary to the sequences of SEQ ID NOs: 38-291.
7. The nucleic acid of Claim 6, wherein said nucleic acid is recombinant.
8. A purified or isolated nucleic acid encoding a human gene product, said human gene product having a sequence partially encoded by one of the sequences of SEQ ID NO: 38-291.
- 20 9. A purified or isolated nucleic acid having the sequence of one of SEQ ID NOs: 38-291 or having a sequence complementary thereto.
10. A purified or isolated nucleic acid comprising the nucleotides of one of SEQ ID NOs: 38-291 which encode a signal peptide.
11. A purified or isolated polypeptides comprising a signal peptide encoded by
25 one of the sequences of SEQ ID NOs: 38-291.
12. A vector encoding a fusion protein comprising a polypeptide and a signal peptide, said vector comprising a first nucleic acid encoding a signal peptide encoded by one of the sequences of SEQ ID NOs: 38-291 operably linked to a second nucleic acid encoding a polypeptide.
- 30 13. A method of directing the extracellular secretion of a polypeptide or the insertion of a polypeptide into the membrane comprising the steps of:

obtaining a vector according to Claim 12; and

introducing said vector into a host cell such that said fusion protein is secreted into the extracellular environment of said host cell or inserted into the membrane of said host cell.

14. A method of importing a polypeptide into a cell comprising contacting said
5 cell with a fusion protein comprising a signal peptide encoded by one of the sequences of SEQ ID NOs: 38-291 operably linked to said polypeptide.

15. A method of making a cDNA encoding a human secretory protein that is partially encoded by one of SEQ ID NOs 38-291, comprising the steps of:

obtaining a cDNA comprising one of the sequences of SEQ ID NOs: 38-291;
10 contacting said cDNA with a detectable probe comprising at least 15 consecutive nucleotides of said sequence of SEQ ID NO: 38-291 or a sequence complementary thereto under conditions which permit said probe to hybridize to said cDNA;

identifying a cDNA which hybridizes to said detectable probe; and

isolating said cDNA which hybridizes to said probe.

16. An isolated or purified cDNA encoding a human secretory protein, said
15 human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-291 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 15.

17. The cDNA of Claim 16 wherein said cDNA comprises the full protein coding
20 sequence partially included in one of the sequences of SEQ ID NOs: 38-291.

18. A method of making a cDNA comprising one of the sequences of SEQ ID
NOs: 38-291, comprising the steps of:

contacting a collection of mRNA molecules from human cells with a first primer
capable of hybridizing to the polyA tail of said mRNA;

25 hybridizing said first primer to said polyA tail;

reverse transcribing said mRNA to make a first cDNA strand;

making a second cDNA strand complementary to said first cDNA strand using at
least one primer comprising at least 15 nucleotides of one of the sequences of SEQ ID NOs
38-291; and

30 isolating the resulting cDNA comprising said first cDNA strand and said second
cDNA strand.

19. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-291 or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 18.

5 20. The cDNA of Claim 19 wherein said cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-291.

21. The method of Claim 18, wherein the second cDNA strand is made by:
contacting said first cDNA strand with a first pair of primers, said first pair of primers comprising a second primer comprising at least 15 consecutive nucleotides of one of the
10 sequences of SEQ ID NOs 38-291 and a third primer having a sequence therein which is included within the sequence of said first primer;

performing a first polymerase chain reaction with said first pair of nested primers to generate a first PCR product;

contacting said first PCR product with a second pair of primers, said second pair of
15 primers comprising a fourth primer, said fourth primer comprising at least 15 consecutive nucleotides of said sequence of one of SEQ ID NO:s 38-291 , and a fifth primer, said fourth and fifth primers being capable of hybridizing to sequences within said first PCR product; and
performing a second polymerase chain reaction, thereby generating a second PCR product.

20 22. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein encoded by one of SEQ ID NOs 38-291, or a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 21.

23. The cDNA of Claim 22 wherein said cDNA comprises the full protein coding
25 sequence partially included in one of the sequences of SEQ ID NOs: 38-291.

24. The method of Claim 18 wherein the second cDNA strand is made by:
contacting said first cDNA strand with a second primer comprising at least 15
consecutive nucleotides of the sequences of SEQ ID NOs: 38-291,
hybridizing said second primer to said first strand cDNA; and
30 extending said hybridized second primer to generate said second cDNA strand.

25. An isolated or purified cDNA encoding a human secretory protein, said human secretory protein comprising the protein partially encoded by one of SEQ ID NOs 38-291 or comprising a fragment thereof of at least 10 amino acids, said cDNA being obtainable by the method of Claim 24.

5 26. The cDNA of Claim 25, wherein said cDNA comprises the full protein coding sequence partially included in one of the sequences of SEQ ID NOs: 38-291.

27. A method of making a protein comprising one of the sequences of SEQ ID NO: 292-545, comprising the steps of:

10 obtaining a cDNA encoding the full protein sequence partially included in one of the sequences of sequence of SEQ ID NO: 38-291;

 inserting said cDNA in an expression vector such that said cDNA is operably linked to a promoter;

 introducing said expression vector into a host cell whereby said host cell produces the protein encoded by said cDNA; and

15 isolating said protein.

28. An isolated protein obtainable by the method of Claim 27.

29. A method of obtaining a promoter DNA comprising the steps of:

 obtaining DNAs located upstream of the nucleic acids of SEQ ID NO: 38-291 or the sequences complementary thereto;

20 screening said upstream DNAs to identify a promoter capable of directing transcription initiation; and

 isolating said DNA comprising said identified promoter.

30. The method of Claim 29, wherein said obtaining step comprises chromosome walking from said nucleic acids of SEQ ID NO: 38-291 or sequences complementary thereto.

25 31. The method of Claim 30, wherein said screening step comprises inserting said upstream sequences into a promoter reporter vector.

32. The method of Claim 30, wherein said screening step comprises identifying motifs in said upstream DNAs which are transcription factor binding sites or transcription start sites.

30 33. An isolated promoter obtainable by the method of Claim 32.

34. An isolated or purified protein comprising one of the sequences of SEQ ID NO: 292-545.

35. In an array of discrete ESTs or fragments thereof of at least 15 nucleotides in length, the improvement comprising inclusion in said array of at least one of the sequences of
5 SEQ ID NOs: 38-291, or one of the sequences complementary to the sequences of SEQ ID NOs: 38-291, or a fragment thereof of at least 15 consecutive nucleotides.

36. The array of Claim 35 including therein at least two of the sequences of SEQ ID NOs: 38-291, the sequences complementary to the sequences of SEQ ID NOs: 38-291, or fragments thereof of at least 15 consecutive nucleotides.

10 37. The array of Claim 35 including therein at least five of the sequences of SEQ ID NOs: 38-291, the sequences complementary to the sequences of SEQ ID NOs: 38-291, or fragments thereof of at least 15 consecutive nucleotides.

1/4

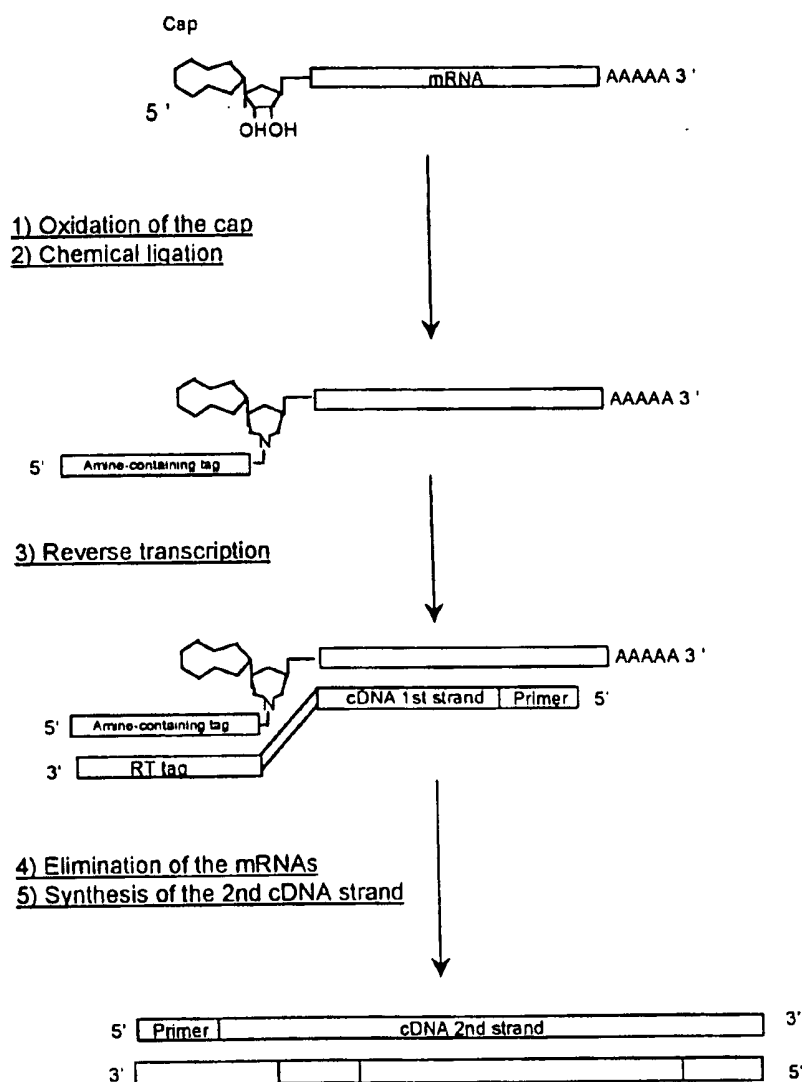


Figure 1

2/4

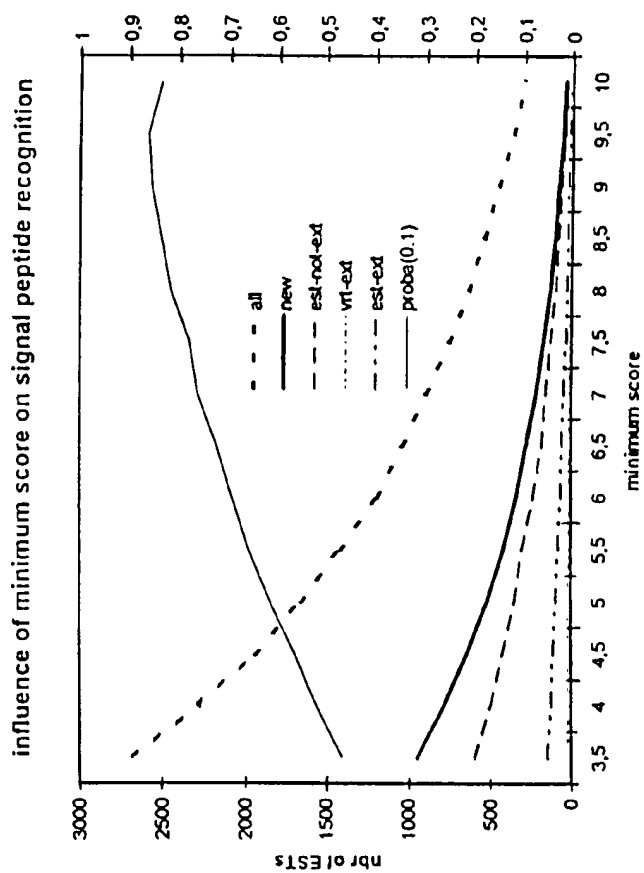


Figure 2

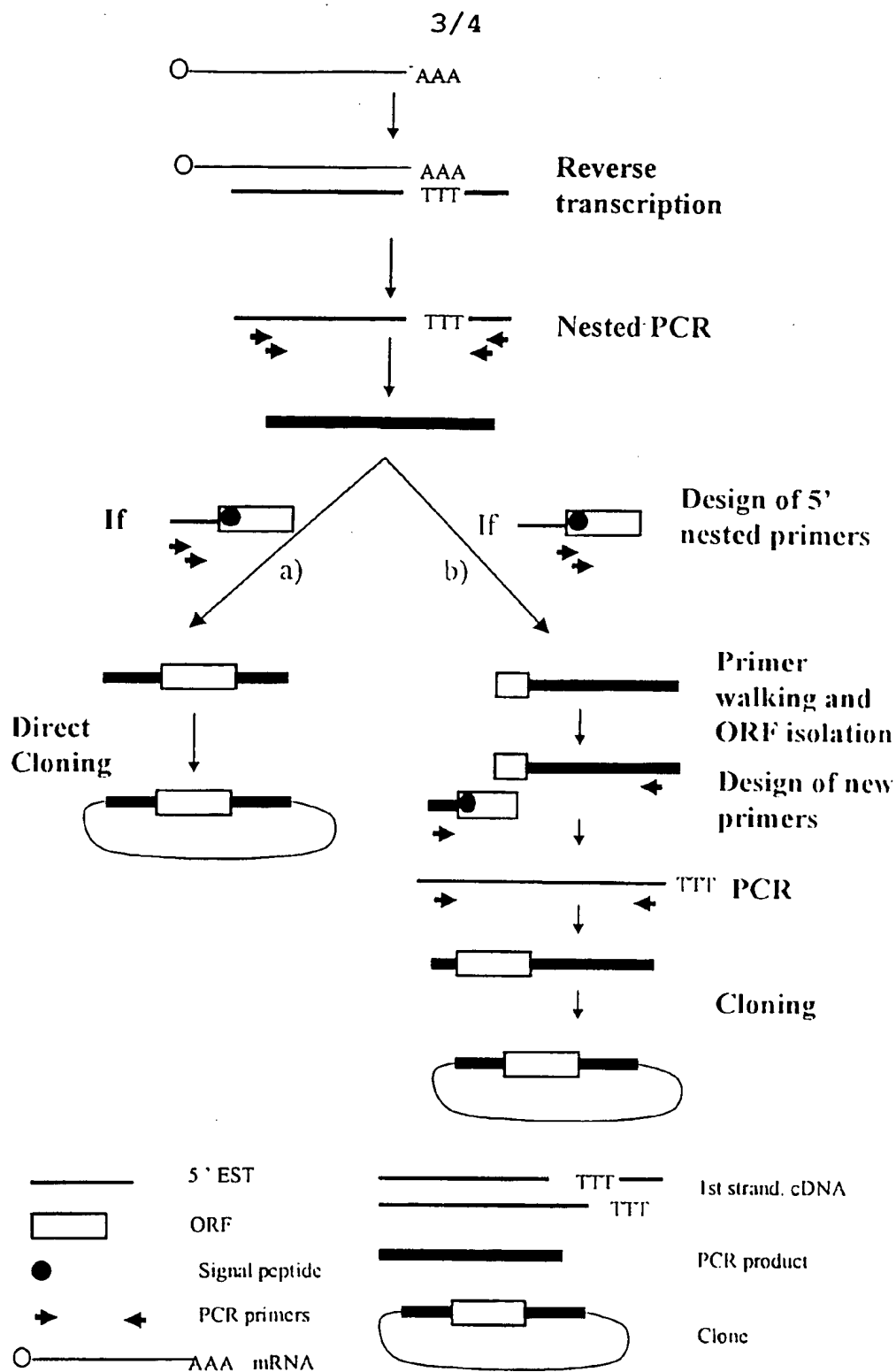
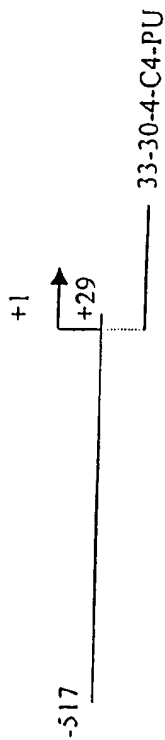
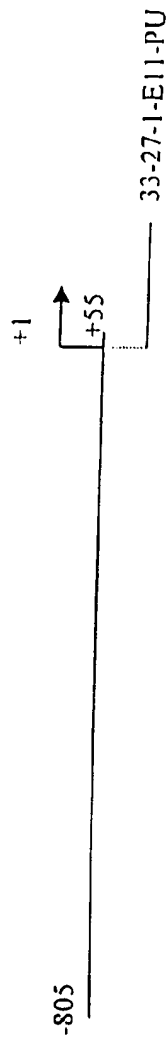


Figure 3

Promoter P13H2



Promoter P15B4



Promoter P29B6

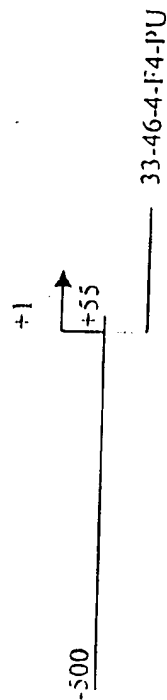


Figure 4

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT:
 - (A) NAME : GENSET SA
 - (B) STREET : 24, RUE ROYALE
 - (C) CITY: PARIS
 - (E) COUNTRY : FRANCE
 - (F) POSTAL CODE (ZIP) : 75008
- (ii) TITLE OF INVENTION: 5' EST FOR NON-TISSUE SPECIFIC
SECRETED PROTEINS
- (iii) NUMBER OF SEQUENCES: 545
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy Disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: Win95
 - (D) SOFTWARE: Word

(2) INFORMATION FOR SEQ ID NO: 1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 47 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: Other nucleic acid
- (ix) FEATURE:
 - (A) NAME/KEY: Cap
 - (B) LOCATION: 1
 - (D) OTHER INFORMATION: m7Gppp added to 1
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGCAUCCUAC UCCCAUCCAA UCCACCCUA ACUCCUCCCA UCUCCAC

47

(2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 46 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: Other nucleic acid
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GCAUCCUACU CCCAUCCAAU UCCACCCUAA CUCCUCCCAU CUCCAC

46

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 base pairs
 (B) TYPE: NUCLEIC ACID
 (C) STRANDEDNESS: SINGLE
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

ATCAAGAATT CGCACGAGAC CATT

25

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 base pairs
 (B) TYPE: NUCLEIC ACID
 (C) STRANDEDNESS: SINGLE
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

TAATGGTCTC GTGCGAATTC TTGAT

25

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 base pairs
 (B) TYPE: NUCLEIC ACID
 (C) STRANDEDNESS: SINGLE
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CCGACAAGAC CAACGTCAAG GCCGC

25

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 base pairs
 (B) TYPE: NUCLEIC ACID
 (C) STRANDEDNESS: SINGLE
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

TCACCAGCAG GCAGTGGCTT AGGAG

25

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

AGTGATTCTT GCTACTTTGG ATGGC

25

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GCTTGGTCTT GTTCTGGAGT TTAGA

25

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

TCCAGAATGG GAGACAAGCC AATTT

25

(2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

AGGGAGGAGG AAACAGCGTG AGTCC

25

(2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

ATGGGAAAGG AAAAGACTCA TATCA

25

(2) INFORMATION FOR SEQ ID NO: 12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

AGCAGCAACA ATCAGGACAG CACAG

25

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

ATCAAGAATT CGCAGGAGAC CATTA

25

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 67 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

ATCGTTGAGA CTCGTACCAG CAGAGTCACG AGAGAGACTA CACGGTACTG GTTTTTTTTT 60

TTTTTVN

67

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CCAGCAGAGT CACGAGAGAG ACTACACGG

29

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

CACGAGAGAG ACTACACGGT ACTGG

25

(2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 526 base pairs
 (B) TYPE: NUCLEIC ACID
 (C) STRANDEDNESS: DOUBLE
 (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo Sapiens
 (F) TISSUE TYPE: Lymph ganglia
- (ix) FEATURE:
 (A) NAME/KEY: other
 (B) LOCATION: complement(261..376)
 (C) IDENTIFICATION METHOD: blastn
 (D) OTHER INFORMATION: identity 96
 region 166..281
 id N70479
 est
- (ix) FEATURE:
 (A) NAME/KEY: other
 (B) LOCATION: complement(380..486)
 (C) IDENTIFICATION METHOD: blastn
 (D) OTHER INFORMATION: identity 97
 region 54..160
 id N70479
 est
- (ix) FEATURE:
 (A) NAME/KEY: other
 (B) LOCATION: complement(110..145)
 (C) IDENTIFICATION METHOD: blastn
 (D) OTHER INFORMATION: identity 94
 region 403..438
 id N70479
 est
- (ix) FEATURE:
 (A) NAME/KEY: other
 (B) LOCATION: complement(196..229)
 (C) IDENTIFICATION METHOD: blastn
 (D) OTHER INFORMATION: identity 94
 region 315..348
 id N70479
 est
- (ix) FEATURE:
 (A) NAME/KEY: sig_peptide
 (B) LOCATION: 90..140
 (C) IDENTIFICATION METHOD: Von Heijne matrix
 (D) OTHER INFORMATION: score 8.2
 seq LLLITAILAVAVG/FP
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

AATATRARAC AGCTACAATA TTCCAGGGCC ARTCACTTGC CATTTCTCAT AACAGCGTCA

60


```

GAGAGAAAGA ACTGACTGAR ACGTTTGAG ATG AAG AAA GTT CTC CTC CTG ATC      113
                               Met Lys Lys Val Leu Leu Leu Ile
                               -15                               -10

ACA GCC ATC TTG GCA GTG GCT GTW GGT TTC CCA GTC TCT CAA GAC CAG      161
Thr Ala Ile Leu Ala Val Ala Val Gly Phe Pro Val Ser Gln Asp Gln
                               -5                               1                               5

GAA CGA GAA AAA AGA AGT ATC AGT GAC AGC GAT GAA TTA GCT TCA GGR      209
Glu Arg Glu Lys Arg Ser Ile Ser Asp Ser Asp Glu Leu Ala Ser Gly
                               10                               15                               20

WTT TTT GTG TTC CCT TAC CCA TAT CCA TTT CGC CCA CTT CCA CCA ATT      257
Xaa Phe Val Phe Pro Tyr Pro Tyr Pro Phe Arg Pro Leu Pro Pro Ile
                               25                               30                               35

CCA TTT CCA AGA TTT CCA TGG TTT AGA CGT AAN TTT CCT ATT CCA ATA      305
Pro Phe Pro Arg Phe Pro Trp Phe Arg Arg Xaa Phe Pro Ile Pro Ile
                               40                               45                               50                               55

CCT GAA TCT GCC CCT ACA ACT CCC CTT CCT AGC GAA AAG TAAACAARAA      354
Pro Glu Ser Ala Pro Thr Thr Pro Leu Pro Ser Glu Lys
                               60                               65

GGAAAAGTCA CRATAAACCT GGTCACCTGA AATTGAAATT GAGCCACTTC CTTGAARAAT      414

CAAAATTCCT GTTAATAAAA RAAAAACAAA TGTAATTGAA ATAGCACACA GCATTCTCTA      474

GTCAATATCT TTAGTGATCT TCTTTAATAA ACATGAAAGC AAAAAAAAAA AA      526

```

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 1..17
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.2
seq LLLITAILAVAVG/FP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

```

Met Lys Lys Val Leu Leu Ile Thr Ala Ile Leu Ala Val Ala Val
 1             5             10             15

Gly

```

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 822 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (D) DEVELOPMENTAL STAGE: Fetal
- (F) TISSUE TYPE: kidney

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 260..464
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96
region 153..357
id H57434
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 118..184
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98
region 98..164
id H57434
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 56..113
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98
region 35..92
id H57434
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 454..485
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100
region 348..379
id H57434
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 118..545
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98
region 1..428
id N27248
est

(ix) FEATURE:

(A) NAME/KEY: other
(B) LOCATION: 65..369
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 98
region 41..345
id H94779
est

(ix) FEATURE:

(A) NAME/KEY: other
(B) LOCATION: 61..399
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 99
region 6..344
id H09880
est

(ix) FEATURE:

(A) NAME/KEY: other
(B) LOCATION: 408..458
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 92
region 355..405
id H09880
est

(ix) FEATURE:

(A) NAME/KEY: other
(B) LOCATION: 60..399
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 97
region 56..395
id H29351
est

(ix) FEATURE:

(A) NAME/KEY: other
(B) LOCATION: 393..432
(C) IDENTIFICATION METHOD: blastn
(D) OTHER INFORMATION: identity 90
region 391..430
id H29351
est

(ix) FEATURE:

(A) NAME/KEY: sig_peptide
(B) LOCATION: 346..408
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 5.5
seq SFLPSALVIWTS/AF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

ACTCCTTTTA GCATAGGGGC TTCGGCGCCA GCGGCCAGCG CTAGTCGGTC TGGTAAGTGC	60
CTGATGCCGA GTTCCGTCTC TCGCGTCTTT TCCTGGTCCC AGGCAAAGCG GASGNAGATC	120
CTCAAACGGC CTAGTGCTTC GCGCTTCCGG AGAAAATCAG CCGTCTAATT AATTCCTCTG	180
GTTTGTTGAA GCAGTTACCA AGAATCTTCA ACCCTTTCCC ACAAAGCTA ATTGAGTACA	240

```

CGTTCCTGTT GAGTACACGT TCCTGTTGAT TTACAAAAGG TGCAGGTATG AGCAGGTCTG      300
AAGACTAACA TTTTGTGAAG TTGTAAAACA GAAAACCTGT TAGAA ATG TGG TGG TTT      357
                               Met Trp Trp Phe
                               -20

CAG CAA GGC CTC AGT TTC CTT CCT TCA GCC CTT GTA ATT TGG ACA TCT      405
Gln Gln Gly Leu Ser Phe Leu Pro Ser Ala Leu Val Ile Trp Thr Ser
   -15                -10                -5

GCT GCT TTC ATA TTT TCA TAC ATT ACT GCA GTA ACA CTC CAC CAT ATA      453
Ala Ala Phe Ile Phe Ser Tyr Ile Thr Ala Val Thr Leu His His Ile
   1                5                10                15

GAC CCG GCT TTA CCT TAT ATC AGT GAC ACT GGT ACA GTA GCT CCA RAA      501
Asp Pro Ala Leu Pro Tyr Ile Ser Asp Thr Gly Thr Val Ala Pro Xaa
           20                25                30

AAA TGC TTA TTT GGG GCA ATG CTA AAT ATT GCG GCA GTT TTA TGT CAA      549
Lys Cys Leu Phe Gly Ala Met Leu Asn Ile Ala Ala Val Leu Cys Gln
           35                40                45

AAA TAGAAATCAG GAARATAATT CAACTTAAAG AAKTTCATTT CATGACCAAA      602
Lys

CTCTTCARAA ACATGTCTTT ACAAGCATAT CTCTTGATT GCTTTCTACA CTGTTGAATT      662
GTCTGGCAAT ATTTCTGCAG TGGAAAATTT GATTTARMTA GTTCTTGACT GATAAATATG      722
GTAAGGTGGG CTTTTCCCCC TGTGTAATTG GCTACTATGT CTTACTGAGC CAAGTTGTAW      782
TTTGAAATAA AATGATATGA GAGTGACACA AAAAAAAAAA      822

```

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 1..21
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.5
seq SFLPSALVIWTS/AF

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

```

Met Trp Trp Phe Gln Gln Gly Leu Ser Phe Leu Pro Ser Ala Leu Val
 1                5                10                15

```

Ile Trp Thr Ser Ala
20

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 405 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Testis

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(103..398)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96
region 1..296
id AA442893
est

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 185..295
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.9
seq LSYASSALSPCLT/AP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

ATCACCTTCT TCTCCATCCT TSTCTGGGCC AGTCCCCARC CCAGTCCCTC TCCTGACCTG	60
CCCAGCCCCAA GTCAGCCTTC AGCACGCGCT TTTCTGCACA CAGATATTCC AGGCCTACCT	120
GGCATTCCAG GACCTCCGMA ATGATGCTCC AGTCCCTTAC AAGCGCTTCC TGGATGAGGG	180
TGGC ATG GTG CTG ACC ACC CTC CCC TTG CCC TCT GCC AAC AGC CCT GTG	229
Met Val Leu Thr Thr Leu Pro Leu Pro Ser Ala Asn Ser Pro Val	
-35 -30 -25	
AAC ATG CCC ACC ACT GGC CCC AAC AGC CTG AGT TAT GCT AGC TCT GCC	277
Asn Met Pro Thr Thr Gly Pro Asn Ser Leu Ser Tyr Ala Ser Ser Ala	
-20 -15 -10	
CTG TCC CCC TGT CTG ACC GCT CCA AAK TCC CCC CGG CTT GCT ATG ATG	325
Leu Ser Pro Cys Leu Thr Ala Pro Xaa Ser Pro Arg Leu Ala Met Met	
-5 1 5 10	
CCT GAC AAC TAAATATCCT TATCCAAATC AATAAARWRA RAATCCTCCC TCCARAAGGG	384
Pro Asp Asn	

TTTCTAAAAA CAAAAAAAAA A

405

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 1..37
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.9
seq LSYASSALSPCLT/AP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

Met Val Leu Thr Thr Leu Pro Leu Pro Ser Ala Asn Ser Pro Val Asn
1 5 10 15
Met Pro Thr Thr Gly Pro Asn Ser Leu Ser Tyr Ala Ser Ser Ala Leu
20 25 30
Ser Pro Cys Leu Thr
35

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 496 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 149..331
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98
region 1..183
id AA397994
est

(ix) FEATURE:

(A) NAME/KEY: other
 (B) LOCATION: 328..485
 (C) IDENTIFICATION METHOD: blastn
 (D) OTHER INFORMATION: identity 96
 region 179..336
 id AA397994
 est

(ix) FEATURE:

(A) NAME/KEY: other
 (B) LOCATION: complement(182..496)
 (C) IDENTIFICATION METHOD: blastn
 (D) OTHER INFORMATION: identity 97
 region 14..328
 id AA399680
 est

(ix) FEATURE:

(A) NAME/KEY: sig_peptide
 (B) LOCATION: 196..240
 (C) IDENTIFICATION METHOD: Von Heijne matrix
 (D) OTHER INFORMATION: score 5.5
 seq ILSTVTALTFAXA/LD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

```

AAAAAATTGG TCCCAGTTTT CACCCTGCCG CAGGGCTGGC TGGGGAGGGC AGCGGTTTAG      60
ATTAGCCGTG GCCTAGGCCG TTAAACGGGG TGACACGAGC NTGCAGGGCC GAGTCCAAGG      120
CCCGGAGATA GGACCAACCG TCAGGAATGC GAGGAATGTT TTTCTCGGA CTCTATCGAG      180
GCACACAGAC AGACC ATG GGG ATT CTG TCT ACA GTG ACA GCC TTA ACA TTT      231
          Met Gly Ile Leu Ser Thr Val Thr Ala Leu Thr Phe
          -15                -10                -5

GCC ARA GCC CTG GAC GGC TGC AGA AAT GGC ATT GCC CAC CCT GCA AGT      279
Ala Xaa Ala Leu Asp Gly Cys Arg Asn Gly Ile Ala His Pro Ala Ser
          1                5                10

GAG AAG CAC AGA CTC GAG AAA TGT AGG GAA CTC GAG ASC ASC CAC TCG      327
Glu Lys His Arg Leu Glu Lys Cys Arg Glu Leu Glu Xaa Xaa His Ser
          15                20                25

GCC CCA GGA TCA ACC CAS CAC CGA AGA AAA ACA ACC AGA AGA AAT TAT      375
Ala Pro Gly Ser Thr Xaa His Arg Arg Lys Thr Thr Arg Arg Asn Tyr
          30                35                40                45

TCT TCA GCC TGAAATGAAK CCGGGATCAA ATGGTTGCTG ATCARAGCCC ATATTAAAT      434
Ser Ser Ala

TGAAAAGTC AAATTGASCA TTATTAAATA AAGCTTGTTT AATATGTCTC AAACAAAAAA      494
AA                                                                 496
  
```

(2) INFORMATION FOR SEQ ID NO: 24:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: AMINO ACID
 (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: PROTEIN
- (vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo Sapiens
- (ix) FEATURE:
 (A) NAME/KEY: sig_peptide
 (B) LOCATION: 1..15
 (C) IDENTIFICATION METHOD: Von Heijne matrix
 (D) OTHER INFORMATION: score 5.5
 seq ILSTVTALTFAXA/LD
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Met Gly Ile Leu Ser Thr Val Thr Ala Leu Thr Phe Ala Xaa Ala
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 623 base pairs
 (B) TYPE: NUCLEIC ACID
 (C) STRANDEDNESS: DOUBLE
 (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo Sapiens
 (F) TISSUE TYPE: Testis
- (ix) FEATURE:
 (A) NAME/KEY: sig_peptide
 (B) LOCATION: 49..96
 (C) IDENTIFICATION METHOD: Von Heijne matrix
 (D) OTHER INFORMATION: score 10.1
 seq LVLTLCTLPLAVA/SA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

AAAGATCCCT GCAGCCCGGC AGGAGAGAAG GCTGAGCCTT CTGGCGTC ATG GAG AGG 57
 Met Glu Arg
 -15

CTC GTC CTA ACC CTG TGC ACC CTC CCG CTG GCT GTG GCG TCT GCT GGC 105
 Leu Val Leu Thr Leu Cys Thr Leu Pro Leu Ala Val Ala Ser Ala Gly
 -10 -5 1

TGC GCC ACG ACG CCA GCT CGC AAC CTG AGC TGC TAC CAG TGC TTC AAG 153
 Cys Ala Thr Thr Pro Ala Arg Asn Leu Ser Cys Tyr Gln Cys Phe Lys
 5 10 15

GTC AGC AGC TGG ACG GAG TGC CCG CCC ACC TGG TGC AGC CCG CTG GAC	201
Val Ser Ser Trp Thr Glu Cys Pro Pro Thr Trp Cys Ser Pro Leu Asp	
20 25 30 35	
CAA GTC TGC ATC TCC AAC GAG GTG GTC GTC TCT TTT AAA TGG AGT GTA	249
Gln Val Cys Ile Ser Asn Glu Val Val Val Ser Phe Lys Trp Ser Val	
40 45 50	
CGC GTC CTG CTC AGC AAA CGC TGT GCT CCC AGA TGT CCC AAC GAC AAC	297
Arg Val Leu Leu Ser Lys Arg Cys Ala Pro Arg Cys Pro Asn Asp Asn	
55 60 65	
ATG AAK TTC GAA TGG TCG CCG GCC CCC ATG GTG CAA GGC GTG ATC ACC	345
Met Xaa Phe Glu Trp Ser Pro Ala Pro Met Val Gln Gly Val Ile Thr	
70 75 80	
AGG CGC TGC TGT TCC TGG GCT CTC TGC AAC AGG GCA CTG ACC CCA CAG	393
Arg Arg Cys Cys Ser Trp Ala Leu Cys Asn Arg Ala Leu Thr Pro Gln	
85 90 95	
GAG GGG CGC TGG GCC CTG CRA GGG GGG CTC CTG CTC CAG GAC CCT TCG	441
Glu Gly Arg Trp Ala Leu Xaa Gly Gly Leu Leu Leu Gln Asp Pro Ser	
100 105 110 115	
AGG GGC ARA AAA ACC TGG GTG CGG CCA CAG CTG GGG CTC CCA CTC TGC	489
Arg Gly Xaa Lys Thr Trp Val Arg Pro Gln Leu Gly Leu Pro Leu Cys	
120 125 130	
CTT CCC AWT TCC AAC CCC CTC TGC CCA RGG GAA ACC CAG GAA GGA	534
Leu Pro Xaa Ser Asn Pro Leu Cys Pro Xaa Glu Thr Gln Glu Gly	
135 140 145	
TAACACTGTG GGTGCCCCCA CCTGTGCATT GGGACCACRA CTTCACCCTC TTGGARACAA	594
TAAACTCTCA TGCCCCCAAA AAAAAAAAAA	623

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 1..16
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 10.1
seq LVLTLCTPLAVA/SA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Met Glu Arg Leu Val Leu Thr Leu Cys Thr Leu Pro Leu Ala Val Ala

1

5

10

15

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 848 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (D) DEVELOPMENTAL STAGE: Fetal
- (F) TISSUE TYPE: kidney

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 32..73
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 10.7
seq LWLLFFLVTAIHA/EL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

```

AACTTTGCCT TGTGTTTTCC ACCCTGAAAG A ATG TTG TGG CTG CTC TTT TTT CTG      55
                               Met Leu Trp Leu Leu Phe Phe Leu
                               -10

GTG ACT GCC ATT CAT GCT GAA CTC TGT CAA CCA GGT GCA GAA AAT GCT      103
Val Thr Ala Ile His Ala Glu Leu Cys Gln Pro Gly Ala Glu Asn Ala
-5                               1                               5                               10

TTT AAA GTG AGA CTT AGT ATC AGA ACA GCT CTG GGA GAT AAA GCA TAT      151
Phe Lys Val Arg Leu Ser Ile Arg Thr Ala Leu Gly Asp Lys Ala Tyr
                               15                               20                               25

GCC TGG GAT ACC AAT GAA GAA TAC CTC TTC AAA GCG ATG GTA GCT TTC      199
Ala Trp Asp Thr Asn Glu Glu Tyr Leu Phe Lys Ala Met Val Ala Phe
                               30                               35                               40

TCC ATG AGA AAA GTT CCC AAC AGA GAA GCA ACA GAA ATT TCC CAT GTC      247
Ser Met Arg Lys Val Pro Asn Arg Glu Ala Thr Glu Ile Ser His Val
                               45                               50                               55

CTA CTT TGC AAT GTA ACC CAG AGG GTA TCA TTC TGG TTT GTG GTT ACA      295
Leu Leu Cys Asn Val Thr Gln Arg Val Ser Phe Trp Phe Val Val Thr
                               60                               65                               70

GAC CCT TCA AAA AAT CAC ACC CTT CCT GCT GTT GAG GTG CAA TCA GCC      343
Asp Pro Ser Lys Asn His Thr Leu Pro Ala Val Glu Val Gln Ser Ala
75                               80                               85                               90

ATA AGA ATG AAC AAG AAC CGG ATC AAC AAT GCC TTC TTT CTA AAT GAC      391
Ile Arg Met Asn Lys Asn Arg Ile Asn Asn Ala Phe Phe Leu Asn Asp
                               95                               100                               105

```

CAA ACT CTG GAA TTT TTA AAA ATC CCT TCC ACA CTT GCA CCA CCC ATG Gln Thr Leu Glu Phe Leu Lys Ile Pro Ser Thr Leu Ala Pro Pro Met 110 115 120	439
GAC CCA TCT GTG GCC ATC TGG ATT ATT ATA TTT GGT GTG ATA TTT TGC Asp Pro Ser Val Pro Ile Trp Ile Ile Ile Phe Gly Val Ile Phe Cys 125 130 135	487
ATC ATC ATA GTT GCA ATT GCA CTA CTG ATT TTA TCA GGG ATC TGG CAA Ile Ile Ile Val Ala Ile Ala Leu Leu Ile Leu Ser Gly Ile Trp Gln 140 145 150	535
CGT ADA ARA AAG AAC AAA GAA CCA TCT GAA GTG GAT GAC GCT GAA RAT Arg Xaa Xaa Lys Asn Lys Glu Pro Ser Glu Val Asp Asp Ala Glu Xaa 155 160 165 170	583
AAK TGT GAA AAC ATG ATC ACA ATT GAA AAT GGC ATC CCC TCT GAT CCC Xaa Cys Glu Asn Met Ile Thr Ile Glu Asn Gly Ile Pro Ser Asp Pro 175 180 185	631
CTG GAC ATG AAG GGA GGG CAT ATT AAT GAT GCC TTC ATG ACA GAG GAT Leu Asp Met Lys Gly Gly His Ile Asn Asp Ala Phe Met Thr Glu Asp 190 195 200	679
GAG AGG CTC ACC CCT CTC TGAAGGGCTG TTGTTCTGCT TCCTCAARAA Glu Arg Leu Thr Pro Leu 205	727
ATTAAACATT TGTTTCTGTG TGA CTGCTGA GCATCCTGAA ATACCAAGAG CAGATCATAT	787
WTTTGTGTTT ACCATTCTTC TTTTGTAAATA AATTTTGAAT GTGCTTGAAA AAAAAAAAAA	847
C	848

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: AMINO ACID
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 1..14
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 10.7
seq LWLLFFLVTAIHA/EL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

Met Leu Trp Leu Leu Phe Phe Leu Val Thr Ala Ile His Ala
1 5 10

(2) INFORMATION FOR SEQ ID NO: 29:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

GGGAAGATGG AGATAGTATT GCCTG

25

(2) INFORMATION FOR SEQ ID NO: 30:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CTGCCATGTA CATGATAGAG AGATTC

26

(2) INFORMATION FOR SEQ ID NO: 31:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 546 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Genomic DNA

(ix) FEATURE:

- (A) NAME/KEY: promoter
- (B) LOCATION: 1..517

(ix) FEATURE:

- (A) NAME/KEY: transcription start site
- (B) LOCATION: 518

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 17..25
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name CMYB_01
score 0.983
sequence TGTCAGTTG

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(18..27)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name MYOD_Q6
score 0.961
sequence CCCAACTGAC

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(75..85)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name S8_01
score 0.960
sequence AATAGAATTAG

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 94..104
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name S8_01
score 0.966
sequence AACTAAATTAG

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(129..139)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name DELTAEF1_01
score 0.960
sequence GCACACCTCAG

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(155..165)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name GATA_C
score 0.964
sequence AGATAAATCCA

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 170..178
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name CMYB_01
score 0.958
sequence CTTCAAGTTG

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 176..189
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name GATA1_02
score 0.959
sequence TTGTAGATAGGACA

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 180..190
- (C) IDENTIFICATION METHOD: matinspector prediction

(D) OTHER INFORMATION: name GATA_C
score 0.953
sequence AGATAGGACAT

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: 284..299
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name TAL1ALPHA47_01
score 0.973
sequence CATAACAGATGGTAAG

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: 284..299
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name TAL1BETAE47_01
score 0.983
sequence CATAACAGATGGTAAG

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: 284..299
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name TAL1BETAITF2_01
score 0.978
sequence CATAACAGATGGTAAG

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(287..296)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name MYOD_Q6
score 0.954
sequence ACCATCTGTT

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(302..314)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name GATA1_04
score 0.953
sequence TCAAGATAAAGTA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: 393..405
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name IK1_01
score 0.963
sequence AGTTGGGAATTCC

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: 393..404
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name IK2_01
score 0.985
sequence AGTTGGGAATTC

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: 396..405
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name CREL_01
score 0.962
sequence TGGAATTCC

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: 423..436
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name GATA1_02
score 0.950
sequence TCAGTGATATGGCA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(478..489)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name SRY_02
score 0.951
sequence TAAAACAAAACA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: 486..493
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name E2F_02
score 0.957
sequence TTTAGCGC

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(514..521)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name MZF1_01
score 0.975
sequence TGAGGGGA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

```
TGAGTGCAGT GTTACATGTC AGTTGGGTTA AGTTTGTAA TGTCATTCAA ATCTTCTATG   60
TCTTGATTTG CCGCTAATT CTATTATTTT TGGAACTAAA TAGTTTGAT GGTTCATTA   120
GTTATTGACT GAGGTGTGCT AATCTCCCAT TATGTGGATT TATCTATTTT TTCAGTTGTA   180
GATAGGACAT TGATAGATAC ATAAGTACCA GGACAAAAGC AGGGAGATCT TTTTCCAAA   240
ATCAGGAGAA AAAAATGACA TCTGGAAAAC CTATAGGGAA AGGCATAACA GATGGTAAGG   300
ATACTTTATC TTGAGTAGGA GAGCCTTCCT GTGGCAACGT GGAGAAGGGA AGAGGTCGTA   360
GAATTGAGGA GTCAGCTCAG TTAGAAGCAG GGAGTTGGGA ATTCCGTTCA TGTGATTTAG   420
CATCAGTGAT ATGGCAAATG TGGGACTAAG GGTAGTGATC AGAGGGTTAA AATTGTGTGT   480
TTTGTTTTAG CGCTGCTGGG GCATCGCCTT GGGTCCCCTC AAACAGATTC CCATGAATCT   540
CTTCAT                                         546
```

(2) INFORMATION FOR SEQ ID NO: 32:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

GTACCAGGGA CTGTGACCAT TGC

23

(2) INFORMATION FOR SEQ ID NO: 33:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: SINGLE
 - (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

CTGTGACCAT TGCTCCCAAG AGAG

24

(2) INFORMATION FOR SEQ ID NO: 34:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 861 base pairs
 - (B) TYPE: NUCLEIC ACID
 - (C) STRANDEDNESS: DOUBLE
 - (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Genomic DNA

(ix) FEATURE:

- (A) NAME/KEY: promoter
- (B) LOCATION: 1..806

(ix) FEATURE:

- (A) NAME/KEY: transcription start site
- (B) LOCATION: 807

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(60..70)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name NFY_Q6
score 0.956

sequence GGACCAATCAT

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 70..77
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name MZF1_01
score 0.962
sequence CCTGGGGA

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 124..132
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name CMYB_01
score 0.994
sequence TGACCGTTG

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(126..134)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name VMYB_02
score 0.985
sequence TCCAACGGT

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 135..143
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name STAT_01
score 0.968
sequence TTCCTGGAA

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(135..143)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name STAT_01
score 0.951
sequence TTCCAGGAA

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(252..259)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name MZF1_01
score 0.956
sequence TTGGGGGA

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 357..368
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name IK2_01
score 0.965
sequence GAATGGGATTTC

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 384..391

(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name MZF1_01
score 0.986
sequence AGAGGGGA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(410..421)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name SRY_02
score 0.955
sequence GAAACAAAACA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: 592..599
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name MZF1_01
score 0.960
sequence GAAGGGGA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: 618..627
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name MYOD_Q6
score 0.981
sequence AGCATCTGCC

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: 632..642
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name DELTAEF1_01
score 0.958
sequence TCCCACCTTCC

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(813..823)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name S8_01
score 0.992
sequence GAGGCAATTAT

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(824..831)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name MZF1_01
score 0.986
sequence AGAGGGGA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

TACTATAGGG CACGCGTGGT CGACGGCCGG GCTGTTCTGG AGCAGAGGGC ATGTCAGTAA 60
TGATTGGTCC CTGGGGAAGG TCTGGCTGGC TCCAGCACAG TGAGGCATTT AGGTATCTCT 120
CGGTGACCGT TGGATTCTCTG GAAGCAGTAG CTGTTCTGTT TGGATCTGGT AGGGACAGGG 180

CTCAGAGGGC TAGGCACGAG GGAAGGTCAG AGGAGAAGGS AGGSARGGCC CAGTGAGARG 240
GGAGCATGCC TTCCCCAAC CCTGGCTTSC YCTTGYMAM AGGGCGKTTY TGGGMACTTR 300
AAYTCAGGGC CCAASCAGAA SCACAGGCC AKTCNTGGCT SMAAGCACA TAGCCTGAAT 360
GGGATTTTCTAG GTTAGNCAGG GTGAGAGGGG AGGCTCTCTG GCTTAGTTTT GTTTTGT TTTT 420
CCAAATCAAG GTAACCTGCT CCCTTCTGCT ACGGGCCTTG GTCTTGGCTT GTCCTCACCC 480
AGTCGGA ACT CCCTACCACT TTCAGGAGAG TGTTTTTAGG CCCGTGGGGC TGTCTGTTC 540
CAAGCAGTGT GAGAACATGG CTGGTAGAGG CTCTAGCTGT GTGCGGGGCC TGAAGGGGAG 600
TGGGTTCTCG CCCAAAGAGC ATCTGCCCCAT TTCCACCTT CCCTTCTCCC ACCAGAAGCT 660
TGCCTGAGCT GTTTGGACAA AAATCCAAAC CCCACTTGGC TACTCTGGCC TGGCTTCAGC 720
TTGGAACCCA ATACCTAGGC TTACAGGCCA TCCTGAGCCA GGGGCCTCTG GAAATTCTCT 780
TCCTGATGGT CCTTTAGGTT TGGGCACAAA ATATAATTGC CTCTCCCCTC TCCATTTTC 840
TCTCTTGGGA GCAATGGTCA C 861

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CTGGGATGGA AGGCACGGTA

20

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: SINGLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

GAGACCACAC AGCTAGACAA

20

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 555 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Genomic DNA

(ix) FEATURE:

- (A) NAME/KEY: promoter
- (B) LOCATION: 1..500

(ix) FEATURE:

- (A) NAME/KEY: transcription start site
- (B) LOCATION: 501

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 191..206
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name ARNT_01
score 0.964
sequence GGA⁶CTCACGTGCTGCT

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 193..204
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name NMYC_01
score 0.965
sequence ACTCACGTGCTG

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 193..204
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name USF_01
score 0.985
sequence ACTCACGTGCTG

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(193..204)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name USF_01
score 0.985
sequence CAGCACGTGAGT

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(193..204)
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name NMYC_01
score 0.956
sequence CAGCACGTGAGT

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: complement(193..204)

(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name MYCMAX_02
score 0.972
sequence CAGCACGTGAGT

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: 195..202
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name USF_C
score 0.997
sequence TCACGTGC

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(195..202)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name USF_C
score 0.991
sequence GCACGTGA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(210..217)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name MZF1_01
score 0.968
sequence CATGGGGA

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: 397..410
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name ELK1_02
score 0.963
sequence CTCTCCGGAAGCCT

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: 400..409
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name CETS1P54_01
score 0.974
sequence TCCGGAAGCC

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(460..470)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name AP1_Q4
score 0.963
sequence AGTGACTGAAC

(ix) FEATURE:

(A) NAME/KEY: TF binding-site
(B) LOCATION: complement(460..470)
(C) IDENTIFICATION METHOD: matinspector prediction
(D) OTHER INFORMATION: name AP1FJ_Q2
score 0.961
sequence AGTGACTGAAC

(ix) FEATURE:

- (A) NAME/KEY: TF binding-site
- (B) LOCATION: 547..555
- (C) IDENTIFICATION METHOD: matinspector prediction
- (D) OTHER INFORMATION: name PADS_C
score 1.000
sequence TGTGGTCTC

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

CTATAGGGCA CGCKTGGTCG ACGGCCCGGG CTGGTCTGGT CTGTKGTGGA GTCGGGTTGA 60
AGGACAGCAT TTGKACATC TGGTCTACTG CACCTTCCCT CTGCCGTGCA CTTGGCCTTT 120
KAWAAGCTCA GCACCGGTGC CCATCACAGG GCCGGCAGCA CACACATCCC ATTACTCAGA 180
AGGAACTGAC GGA CTACAGT GCTGCTCCGT CCCCATGAGC TCAGTGGACC TGTCTATGTA 240
GAGCAGTCAG ACAGTGCCTG GGATAGAGTG AGAGTTCAGC CAGTAAATCC AAGTGATTGT 300
CATTCCTGTC TGCATTAGTA ACTCCCAACC TAGATGTGAA AACTTAGTTC TTTCTCATAG 360
GTTGCTCTGC CCATGGTCCC ACTGCAGACC CAGGCACTCT CCGGAAGCCT GGAAATCACC 420
CGTGTCTTCT GCCTGCTCCC GCTCACATCC CACACTTGTG TTCAGTCACT GAGTTACAGA 480
TTTTGCCTCC TCAATTCTC TTGTCTTAGT CCCATCCTCT GTTCCCCTGG CCAGTTTGTC 540
TAGCTGTGTG GTCTC 555

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 231 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Liver

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 25..129
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 15
seq LFLLLLLAASAWG/VT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

AAGAAGCAAA AGAGCAGAGC TACC ATG TCC TCT TGG AGC AGA CAG CGA CCA 51
Met Ser Ser Trp Ser Arg Gln Arg Pro
-35 -30

```

AAA AGC CCA GGG GGC ATT CAA CCC CAT GTT TCT AGA ACT CTG TTC CTG      99
Lys Ser Pro Gly Gly Ile Gln Pro His Val Ser Arg Thr Leu Phe Leu
-25                      -20                      -15

CTG CTG CTG TTG GCA GCC TCA GCC TGG GGG GTC ACC CTG AGC CCC AAA      147
Leu Leu Leu Leu Ala Ala Ser Ala Trp Gly Val Thr Leu Ser Pro Lys
-10                      -5                      1                      5

GAC TGC CAG GTG TTC CGC TCA GAC CAT GGC AGC TCC ATC TCC TGT CAA      195
Asp Cys Gln Val Phe Arg Ser Asp His Gly Ser Ser Ile Ser Cys Gln
10                      15                      20

CCA CCT GCC GAA ATC CCC GGC TAC CTG CCA GCC ACG                      231
Pro Pro Ala Glu Ile Pro Gly Tyr Leu Pro Ala Thr
25                      30

```

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 384 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 97..159
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 13.2
seq LLLXAVLLSLASA/SS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

```

AKGAAGAGCA GCGGCGAGGC GCGGGTGGTG GCTGADTCCG TGGTGGCAGA GGCGAAGGCG      60
ACAGCTCTAG GGGTTGGCAC CGGCCCGAG AGGAGG ATG CGG GTC CGG ATA GGG      114
Met Arg Val Arg Ile Gly
-20

CTG ACG CTG CTG CTG TRT GCG GTG CTG CTG AGC TTG GCC TCG GCG TCC      162
Leu Thr Leu Leu Leu Xaa Ala Val Leu Leu Ser Leu Ala Ser Ala Ser
-15                      -10                      -5                      1

TCG GAT GAA GAA GGC AGC CAG GAT GAA TCC TTA GAT TCC AAG ACT ACT      210
Ser Asp Glu Glu Gly Ser Gln Asp Glu Ser Leu Asp Ser Lys Thr Thr
5                      10                      15

TTG ACA TCA GAT GAG TCA GTA AAG GAC CAT ACT ACT GCA GGC AGA GTA      258
Leu Thr Ser Asp Glu Ser Val Lys Asp His Thr Thr Ala Gly Arg Val
20                      25                      30

GTT GCT GGT CAA ATA TTT CTT GAT TCA GAA GAA TCT GAA TTA GAA TNC      306

```

```

Val Ala Gly Gln Ile Phe Leu Asp Ser Glu Glu Ser Glu Leu Glu Xaa
   35                               40                               45

TCT ATT CAA GAA GAG GAA GAC AGC CTC AAG AGC CAA GAG GGG GAA AGT   354
Ser Ile Gln Glu Glu Glu Asp Ser Leu Lys Ser Gln Glu Gly Glu Ser
   50                               55                               60                               65

GTC ACA GAA GAT ATC AGC TTT CTA GAG TCT   384
Val Thr Glu Asp Ile Ser Phe Leu Glu Ser
           70                               75

```

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 438 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Substantia nigra

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 64..126
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 13.1
seq CVLLLLLLLLLTRS/SE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

```

AATTTTGGAG AGTTAAACT GTGCCTAACA GAGGTGTCCT CTGACTTTTC TTCTGCAAGC   60

TCC ATG TTT TCA CAT CTT CCC TTT GAC TGT GTC CTG CTG CTG CTG CTG   108
Met Phe Ser His Leu Pro Phe Asp Cys Val Leu Leu Leu Leu Leu
   -20                               -15                               -10

CTA CTA CTT ACA AGG TCC TCA GAA GTG GAA TAM ARA GCG GAG GTC GGT   156
Leu Leu Leu Thr Arg Ser Ser Glu Val Glu Xaa Xaa Ala Glu Val Gly
   -5                               1                               5                               10

CAG AAT GCC TAT CTG CCC TGC TTC TAC ACC CCA GCC GCC CCA GGG AAC   204
Gln Asn Ala Tyr Leu Pro Cys Phe Tyr Thr Pro Ala Ala Pro Gly Asn
           15                               20                               25

CTC GTG CCC GTC TGC TGG GGC AAA GGA GCC TGT CCT GTG TTT GAA TGT   252
Leu Val Pro Val Cys Trp Gly Lys Gly Ala Cys Pro Val Phe Glu Cys
           30                               35                               40

GGC AAC GTG GTG CTC AGG ACT GAT GAA AGG GAT GTG AAT TAT TGG ACA   300
Gly Asn Val Val Leu Arg Thr Asp Glu Arg Asp Val Asn Tyr Trp Thr
           45                               50                               55

TCC AGA TAC TGG CTA AAT GGG GAT TTC CGC AAA GGA GAT GTG TCC CTG   348
Ser Arg Tyr Trp Leu Asn Gly Asp Phe Arg Lys Gly Asp Val Ser Leu

```


60	65	70	
ACC ATA GAG AAT GTG ACT CTA GCA GAC AGT GGG ATC TAC TGC TGC CGG			396
Thr Ile Glu Asn Val Thr Leu Ala Asp Ser Gly Ile Tyr Cys Cys Arg			
75	80	85	90
ATC CAA ATC CCA GGC ATA ATG AAT GAT GAA AAA TTT AAC CTG			438
Ile Gln Ile Pro Gly Ile Met Asn Asp Glu Lys Phe Asn Leu			
95	100		

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 145 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (D) DEVELOPMENTAL STAGE: Fetal
- (F) TISSUE TYPE: kidney

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 59..121
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 11.6
seq LLFLFLAVDEAWA/GM

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

AACACTACCT TCCCGAAGTT GAAGGCAAGC GGTGATTGTT TGTAGACGGC GCTTTGTC	58
ATG GGA CCT GTG CGG TTG GGA ATA TTG CTT TTC CTT TTT TTG GCC GTG	106
Met Gly Pro Val Arg Leu Gly Ile Leu Leu Phe Leu Phe Leu Ala Val	
-20 -15 -10	
GAC GAG GCT TGG GCT GGG ATG TTG AAG GAG GAG GGA CGG	145
Asp Glu Ala Trp Ala Gly Met Leu Lys Glu Glu Gly Arg	
-5 1 5	

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 258 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Kidney

(ix) FEATURE:

(A) NAME/KEY: other
 (B) LOCATION: 58..194
 (C) IDENTIFICATION METHOD: blastn
 (D) OTHER INFORMATION: identity 100
 region 44..180
 id AA280744
 est

(ix) FEATURE:

(A) NAME/KEY: sig_peptide
 (B) LOCATION: 25..75
 (C) IDENTIFICATION METHOD: Von Heijne matrix
 (D) OTHER INFORMATION: score 10.7
 seq SLLLAVALGLATA/VS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

AATGGCTGAG GAGGTCGCAG CGCC ATG AAG TCC CTG TCT CTG CTC CTC GCT	51
Met Lys Ser Leu Ser Leu Leu Ala	
-15 -10	
GTG GCT TTG GGC CTG GCG ACC GCC GTC TCA GCA GGA CCC GCG GTG ATC	99
Val Ala Leu Gly Leu Ala Thr Ala Val Ser Ala Gly Pro Ala Val Ile	
-5 1 5	
GAG TGT TGG TTC GTG GAG GAT GCG AGC GGA AAG GGC CTG GCC AAG AGA	147
Glu Cys Trp Phe Val Glu Asp Ala Ser Gly Lys Gly Leu Ala Lys Arg	
10 15 20	
CCC GGT GCA CTG CTG TTG CGC CAG GGA CCG GGG GAA CCG CCG CCC CGG	195
Pro Gly Ala Leu Leu Leu Arg Gln Gly Pro Gly Glu Pro Pro Pro Arg	
25 30 35 40	
CCG GAC CTC GAC CCT GAG CTC TAT CTC AGT GTA CAC GAC CCC GCG GGC	243
Pro Asp Leu Asp Pro Glu Leu Tyr Leu Ser Val His Asp Pro Ala Gly	
45 50 55	
GCC CTC CAG GCT CGG	258
Ala Leu Gln Ala Arg	
60	

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 458 base pairs
 (B) TYPE: NUCLEIC ACID
 (C) STRANDEDNESS: DOUBLE
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens
 (F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 144..191
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 9.6
seq LLTLXLLGGPTWA/GK

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

```

GTTCCCCTGG CGGCCCTCG CTTCTTCCTT CTGGATGGGG GCCCAGGGGG CCCAGGAGAG   60
TATAAAGGCG ATGTGGAGGG TGCCCGGCAC AACAGACGC CCAGTCACAG GCGAGAGCCT   120
GGGATGGCAC CCGGCCAGAG GCC ATG CTG CTG CTC ACG CTT GNH CTC CTG   173
          Met Leu Leu Leu Thr Leu Xaa Leu Leu
          -15                      -10

GGG GGC CCC ACC TGG GCA GGG AAG ATG TAT GGC CCT GGA GGA GGC AAG   221
Gly Gly Pro Thr Trp Ala Gly Lys Met Tyr Gly Pro Gly Gly Gly Lys
   -5                      1                      5                      10

TAT TTC AGC ACC ACT GAA GAC TAC GAC CAT GAA ATC ACA GGG CTG CGG   269
Tyr Phe Ser Thr Thr Glu Asp Tyr Asp His Glu Ile Thr Gly Leu Arg
          15                      20                      25

GTG TCT GTA GGT CTT CTC CTG GTG AAA AGT GTC CAG GTG AAA CTT GGA   317
Val Ser Val Gly Leu Leu Leu Val Lys Ser Val Gln Val Lys Leu Gly
          30                      35                      40

GAC TCC TGG GAC GTG AAA CTG GGA GCC TTA RGT GGG AAT ACC CAG GAA   365
Asp Ser Trp Asp Val Lys Leu Gly Ala Leu Xaa Gly Asn Thr Gln Glu
          45                      50                      55

GTC ASW STG CAG CCA GGC GAA TAC ATC ACA AAA GTC TTT GTC GCC TTC   413
Val Xaa Xaa Gln Pro Gly Glu Tyr Ile Thr Lys Val Phe Val Ala Phe
          60                      65                      70

CAA GCT TTC CTC CGG GGT ATG GTC ATG TAC ACC AGC AAG GAC CGA   458
Gln Ala Phe Leu Arg Gly Met Val Met Tyr Thr Ser Lys Asp Arg
          75                      80                      85

```

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 339 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 109..246

(C) IDENTIFICATION METHOD: Von Heijne matrix
 (D) OTHER INFORMATION: score 9.4
 seq LIILIXIWIWCLG/SQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

```

AATTAATCAC GGAGTTCAG GGAGAAGGAA CTTGTGAAAT GGGGGAGCCG GCTGGGGTTG    60
CCGGCACCAT GGAGTCACCT TTTAGCCCGG GACTCTTTCA CAGGCTGG ATG AAG ATT    117
                                   Met Lys Ile
                                   -45

GGG ATT CTG CTC TCT TTG CTG AAC TCG GTT ATT TCA CAG ACA CTG ATG    165
Gly Ile Leu Leu Ser Leu Leu Asn Ser Val Ile Ser Gln Thr Leu Met
      -40                      -35                      -30

AGC TGC AAT TGG AAG CAG CAA ATG AGA CGT ATG AAA ACA ATT TTG ATA    213
Ser Cys Asn Trp Lys Gln Gln Met Arg Arg Met Lys Thr Ile Leu Ile
      -25                      -20                      -15

ATC TTG ATT KTG ATT TGG ATT TGG TGC CTT GGG AGT CAG ACA TTT GGG    261
Ile Leu Ile Xaa Ile Trp Ile Trp Cys Leu Gly Ser Gln Thr Phe Gly
      -10                      -5                      1                      5

ACA TCA ACA ACC AAA TCT GTA CAG TTA AAG ATA TTA AGG CAG AAC CTC    309
Thr Ser Thr Thr Lys Ser Val Gln Leu Lys Ile Leu Arg Gln Asn Leu
              10                      15                      20

AGC CAC TTT CTC CAG CCT CCT CAA GTT ATT    339
Ser His Phe Leu Gln Pro Pro Gln Val Ile
      25                      30
  
```

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 396 base pairs
 (B) TYPE: NUCLEIC ACID
 (C) STRANDEDNESS: DOUBLE
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: Homo Sapiens
 (F) TISSUE TYPE: Brain

(ix) FEATURE:
 (A) NAME/KEY: sig_peptide
 (B) LOCATION: 115..204
 (C) IDENTIFICATION METHOD: Von Heijne matrix
 (D) OTHER INFORMATION: score 9.4
 seq LPFLLSLFPGALP/VQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

```

AAGTATAGAT TCGGAAACAG AAAACAAAAG CAGGAAAAGT GACCTTAGCC CGGATTCTGC    60
  
```

```

CATCCCCGGA AGGCTTATTC CTCCTATGGG CAAAGGAGCA AAGGGAGCCA GAAG ATG      117
                                     Met
                                     -30

AAA GCG AGC TCA GGG AGG TGC GGG CTG GTG CGG TGG CTG CAG GTA CTG      165
Lys Ala Ser Ser Gly Arg Cys Gly Leu Val Arg Trp Leu Gln Val Leu
      -25                      -20                      -15

TTG CCC TTC CTG TTG TCT TTG TTC CCC GGG GCT CTC CCA GTC CAG ATC      213
Leu Pro Phe Leu Leu Ser Leu Phe Pro Gly Ala Leu Pro Val Gln Ile
      -10                      -5                      1

CGC TAT TCA ATT CCA GAG GAG CTG GCC AAA AAC TCG GTC GTA GGA AAC      261
Arg Tyr Ser Ile Pro Glu Glu Leu Ala Lys Asn Ser Val Val Gly Asn
      5                      10                      15

CTC GCC AAG GAT CTG GGG CTC AGC GTC CGG GAC TTG CCA GCC CGG AAG      309
Leu Ala Lys Asp Leu Gly Leu Ser Val Arg Asp Leu Pro Ala Arg Lys
      20                      25                      30                      35

CTG CGG GTT AGC GCG GAG AAG GAA TAT TTC ACA GTA AAC CCA GAA AGC      357
Leu Arg Val Ser Ala Glu Lys Glu Tyr Phe Thr Val Asn Pro Glu Ser
      40                      45                      50

GGA GAC TTA CTT GTG AGT GAC AGA ATA GAC CGA GAC GTG      396
Gly Asp Leu Leu Val Ser Asp Arg Ile Asp Arg Asp Val
      55                      60

```

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 419 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Normal prostate

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 258..356
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 9.1
seq IIFLCHLLRGLHA/XT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

```

AGTTTTCGGT CGGCCCGGGT GTTCTGCAAG CTGGTCAAAA AGGGGAAGCG GCCCAGATAT      60
GTTAAGTTCT ATGGCCGCTG CAGGGTCTGT GAAGGCGGCG TTGCAGGTGG CCGAGGTGCT      120
CGAAGCCATC GTGAGCTGCT GCGTGGGGGC CCGAGGGACG GCAAGTTTGT TGTACGAAGC      180
CCTACTGGCGA GGTGCTTCTC AGCCGGAATG GAGGCCGCCT CCTGGAGGCG CTACACNKAG      240

```

```

AGCATCCCAT AGCCAGG ATG ATA GTG GAC TGT GTT TCC AGT CAT CTC AAA      290
      Met Ile Val Asp Cys Val Ser Ser His Leu Lys
      -30                               -25

AAA ACA GGA GAT GGT GCA AAA ACA TTT ATT ATC TTT CTT TGC CAT TTG      338
Lys Thr Gly Asp Gly Ala Lys Thr Phe Ile Ile Phe Leu Cys His Leu
      -20                               -15                               -10

CTT AGA GGA CTT CAT GCD MTC ACA GAC AGA GAA AAG GAT CCT TTG ATG      386
Leu Arg Gly Leu His Ala Xaa Thr Asp Arg Glu Lys Asp Pro Leu Met
      -5                               1                               5                               10

TGT GAA AAC ATT CAA ACC CAT GGA AGG CTT CCG                          419
Cys Glu Asn Ile Gln Thr His Gly Arg Leu Pro
      15                               20

```

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 380 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Ovary

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 54..365
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 9.1
seq LTSLSWLLXASCS/KP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

```

AATTGCGCGC CGGCCTCAAG ATGGCCGCCT TCTGGCGTCT CCGGCGCTGT TGA ATG      56
                                     Met

GCG AAA GCT TTA TTG TTC CCT TCG GGC AGG AGT GTT CGT GTC CTC TAT      104
Ala Lys Ala Leu Leu Phe Pro Ser Gly Arg Ser Val Arg Val Leu Tyr
      -100                               -95                               -90

GGC GCT GTC AAT AAA GAA CGG CAG TDT GAA TCG GTG CTG AAC AGG GCC      152
Gly Ala Val Asn Lys Glu Arg Gln Xaa Glu Ser Val Leu Asn Arg Ala
      -85                               -80                               -75

TGT CCT CCC AAA GCC AAC TCT AAG GAG AGG AGA GGA AGA GCA GTT CTT      200
Cys Pro Pro Lys Ala Asn Ser Lys Glu Arg Arg Gly Arg Ala Val Leu
      -70                               -65                               -60

GGG GCA GAG TTG ACG CAA TGG AGC TCC CCA ACT ACA GCC GGC AGC TGC      248
Gly Ala Glu Leu Thr Gln Trp Ser Ser Pro Thr Thr Ala Gly Ser Cys
      -55                               -50                               -45                               -40

```

TGC AGC AGC TGT ACA CTC TGT GCA AGG AGC AGC AGT KCT GTG ATT GCA	296
Cys Ser Ser Cys Thr Leu Cys Ala Arg Ser Ser Ser Xaa Val Ile Ala	
-35 -30 -25	
CCA TCT CCA TTG GTA CCA TTT ACT TCA GGG CTC ACA AGC TTG TCC TGG	344
Pro Ser Pro Leu Val Pro Phe Thr Ser Gly Leu Thr Ser Leu Ser Trp	
-20 -15 -10	
CTG CTG MCA GCM TCC TGT TCA AAA CCC TGM AAA GGG	380
Leu Leu Xaa Ala Ser Cys Ser Lys Pro Xaa Lys Gly	
-5 1 5	

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 428 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Brain

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 27..245
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8
seq LATKLLSLSGVFA/VH

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

AAGAAACAGG TCTGGGCTAC AAAAGT ATG GCC GCT TCT GAG GCG GCG GTG GTG	53
Met Ala Ala Ser Glu Ala Ala Val Val	
-70 -65	
TCT TCG CCG TCT TTG AAA ACA GAC ACA TCC CCT GTC CTT GAA ACT GCA	101
Ser Ser Pro Ser Leu Lys Thr Asp Thr Ser Pro Val Leu Glu Thr Ala	
-60 -55 -50	
GGA ACG GTC GCA GCA ATG GCT GCG ACC CCG TCA GCA AGG GCT GCA GCC	149
Gly Thr Val Ala Ala Met Ala Ala Thr Pro Ser Ala Arg Ala Ala Ala	
-45 -40 -35	
GCG GTG GTT GCG GCC GCG GCC AGG ACC GGA TCC GAA GCC AGG GTC TCC	197
Ala Val Val Ala Ala Ala Ala Arg Thr Gly Ser Glu Ala Arg Val Ser	
-30 -25 -20	
AAG GCC GCT TTG GCT ACC AAG CTG CTG TCC TTG AGC GGC GTG TTC GCC	245
Lys Ala Ala Leu Ala Thr Lys Leu Leu Ser Leu Ser Gly Val Phe Ala	
-15 -10 -5	
GTG CAC AAG CCC AAA GGG CCC ACT TCA GCC GAG CTG CTG AAT CGG TTG	293
Val His Lys Pro Lys Gly Pro Thr Ser Ala Glu Leu Leu Asn Arg Leu	

1	5	10	15	
AAG GAG AAG CTG CTG GCA GAA GCT GGA ATG CCT TCT CCA GAA TGG ACA				341
Lys Glu Lys Leu Leu Ala Glu Ala Gly Met Pro Ser Pro Glu Trp Thr				
	20	25	30	
NAG AGG AAA AAG CAG ACK NHW GAA AAT TGG GCA TGG AGG GAC TCT AGA				389
Xaa Arg Lys Lys Gln Thr Xaa Glu Asn Trp Ala Trp Arg Asp Ser Arg				
	35	40	45	
CAG CGC ASC CGA GGA GTT CTG GTT GTT GGA ATT GGA GCG				428
Gln Arg Xaa Arg Gly Val Leu Val Val Gly Ile Gly Ala				
	50	55	60	

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 332 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (D) DEVELOPMENTAL STAGE: Fetal
- (F) TISSUE TYPE: kidney

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
- (B) LOCATION: 201..251
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 7.8
seq VLWLISFFTFDTG/HG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

AATTGCTGAT GGATCAGTGA GCCTGTGTTC ATGCCAGTGA GCTGCTGTGG CTCAGATACT	60
GATACTTTCT TTCCAAACAG CATAAGAAGT GATTGANCCA CAAGTATACT GAAGGMARGG	120
YHCCCWSVAR TYCTGGWGTG AMGAGATAAA TCACCACTCA CAGACTATGC ACCCGACTGC	180
TGCTGTTCAG TCCAGGGAAA ATG AAA GTT GGA GTG CTG TGG CTC ATT TCT TTC	233
Met Lys Val Gly Val Leu Trp Leu Ile Ser Phe	
-15 -10	
TTC ACC TTC ACT GAC GGC CAC GGT GGC TTC CTG GGG GTG AGT TGG TGC	281
Phe Thr Phe Thr Asp Gly His Gly Gly Phe Leu Gly Val Ser Trp Cys	
-5 1 5 10	
TAT GTC TCA TAT CTC TTC TCA ACT AAC TCT CCT CTC TCG TTC CGG CGC	329
Tyr Val Ser Tyr Leu Phe Ser Thr Asn Ser Pro Leu Ser Phe Arg Arg	
15 20 25	
ATG	332
Met	

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 437 base pairs
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: DOUBLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
(F) TISSUE TYPE: S adrenals

(ix) FEATURE:

- ```
(A) NAME/KEY: sig_peptide
(B) LOCATION: 81..137
(C) IDENTIFICATION METHOD: Von Heijne matrix
(D) OTHER INFORMATION: score 7.4
 seq WIFLAAILKGVOC/EV
```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

|             |             |             |             |             |            |     |
|-------------|-------------|-------------|-------------|-------------|------------|-----|
| AGCTCTGGGA  | GAGGAGCCCC  | AGCCTTGGGA  | TTCCCAAGTG  | TTTTTCATTCA | GTGAGCAGGA | 60  |
| CTGAACACAG  | AGGACTCACC  | ATG GAG TTT | GGG CTR AGC | TGG ATT TTC | CTT GCA    | 113 |
|             |             | Met Glu Phe | Gly Leu Ser | Trp Ile Phe | Leu Ala    |     |
|             |             |             | -15         |             | -10        |     |
| GCT ATT TTA | AAA GGT GTC | CAG TGT GAG | GTG CAG CTG | GTG GAG TCT | GGG        | 161 |
| Ala Ile Leu | Lys Gly Val | Gln Cys Glu | Val Gln Leu | Val Glu Ser | Gly        |     |
|             | -5          |             | 1           | 5           |            |     |
| GGA GGC TTG | GTA AAG CCT | GGG GGG TCC | CTG AGA CTC | TCC TGT GCA | GCC        | 209 |
| Gly Gly Leu | Val Lys Pro | Gly Gly Ser | Leu Arg Leu | Ser Cys Ala | Ala        |     |
|             | 10          | 15          |             | 20          |            |     |
| TCT GGA TTC | GAT TTC ACT | GAC GCC TGG | ATG AGT TGG | GTC CGC CAG | GCT        | 257 |
| Ser Gly Phe | Asp Phe Thr | Asp Ala Trp | Met Ser Trp | Val Arg Gln | Ala        |     |
|             | 25          | 30          | 35          |             | 40         |     |
| CCG GGG AAG | GGG CTG GAG | TGG GTT GCC | AAT ATA NGA | AGC ACA GCC | TCT        | 305 |
| Pro Gly Lys | Gly Leu Glu | Trp Val Ala | Asn Ile Xaa | Ser Thr Ala | Ser        |     |
|             | 45          |             | 50          |             | 55         |     |
| GGT GGG ACA | AGA GGC TAC | GCT GCA CCC | GTG AAA GAC | AGA TTC ATC | ATC        | 353 |
| Gly Gly Thr | Arg Gly Tyr | Ala Ala Pro | Val Lys Asp | Arg Phe Ile | Ile        |     |
|             | 60          | 65          |             | 70          |            |     |
| TCA AGG GAT | GAT TCA AGA | AAC ACT CTA | CAC CTA CAA | ATG AAC GGC | CTG        | 401 |
| Ser Arg Asp | Asp Ser Arg | Asn Thr Leu | His Leu Gln | Met Asn Gly | Leu        |     |
|             | 75          | 80          |             | 85          |            |     |
| AAA MCG ATG | ACA CAR GCC | ATC TAT TAT | TGT GCC ACA |             |            | 437 |
| Lys Xaa Met | Thr Gln Ala | Ile Tyr Tyr | Cys Ala Thr |             |            |     |
|             | 90          | 95          | 100         |             |            |     |

## (2) INFORMATION FOR SEQ ID NO: 51:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 466 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Cancerous prostate

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 17..127
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 7.4  
seq LWRLLLWAGTAFQ/VX

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

```

AACTCAGGAC AACGCT ATG GCT GAG CCT GGG CAC AGC CAC CAT CTC TCC GCC 52
 Met Ala Glu Pro Gly His Ser His His Leu Ser Ala
 -35 -30

AGA GTC AGG GGA AGA ACT GAG AGG CGC ATA CCC CGG CTG TGG CGG CTG 100
Arg Val Arg Gly Arg Thr Glu Arg Arg Ile Pro Arg Leu Trp Arg Leu
-25 -20 -15 -10

CTG CTC TGG GCT GGG ACC GCC TTC CAG GTG RMC CAG GGA MSG GRA CCG 148
Leu Leu Trp Ala Gly Thr Ala Phe Gln Val Xaa Gln Gly Xaa Xaa Pro
 -5 1 5

GAG CTT CAS GCC TGC AAA GAG TCT GAG TAC CAC TAT GAG TAC ACG GCG 196
Glu Leu Xaa Ala Cys Lys Glu Ser Glu Tyr His Tyr Glu Tyr Thr Ala
 10 15 20

TGT GAC AGC ACG GGT TCC AGG TGG AGG GTC GCC GTG CCG CAT ACH YCG 244
Cys Asp Ser Thr Gly Ser Arg Trp Arg Val Ala Val Pro His Thr Xaa
 25 30 35

GGC CTG TGC ACC AGC CTG CCT GAC CCC GTC AAG GGC ACC GAG TGC TSN 292
Gly Leu Cys Thr Ser Leu Pro Asp Pro Val Lys Gly Thr Glu Cys Xaa
 40 45 50 55

NTC TCC TGC AAC GCC GGG GAG TTT CTG GAT ATG AAG GAC CAG TCA TGT 340
Xaa Ser Cys Asn Ala Gly Glu Phe Leu Asp Met Lys Asp Gln Ser Cys
 60 65 70

NNG CCA TGC GCT GAG GGC CGC TAC TCC CTC GGC ACA GGC ATT CGG TTT 388
Xaa Pro Cys Ala Glu Gly Arg Tyr Ser Leu Gly Thr Gly Ile Arg Phe
 75 80 85

GAT GAG TGG GAT GAG CTG CCC CAT GGC TTT GCA GGC TCT CAG CCA ACA 436
Asp Glu Trp Asp Glu Leu Pro His Gly Phe Ala Ala Ser Gln Pro Thr
 90 95 100

```

TGG AGC TGG ATG ACA GTG CTG CTG AGT CAC  
 Trp Ser Trp Met Thr Val Leu Leu Ser His  
 105 110

466

## (2) INFORMATION FOR SEQ ID NO: 52:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 318 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 4..78
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 7.1  
seq QACLLGLFALILS/GK

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

|                                                                                                                                                   |     |
|---------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| AAC ATG ACA GCA GAT CCG CGG AAG GGC AGA ATG GGA CTC CAA GCC TGC<br>Met Thr Ala Asp Pro Arg Lys Gly Arg Met Gly Leu Gln Ala Cys<br>-25 -20 -15     | 48  |
| CTC CTA GGG CTC TTT GCC CTC ATC CTC TCT GGC AAA TGC AGT BAC AGC<br>Leu Leu Gly Leu Phe Ala Leu Ile Leu Ser Gly Lys Cys Ser Xaa Ser<br>-10 -5 1 5  | 96  |
| CCG GAG CCC GAC CAG CGG AGG ACG CTG CCC CCA GGC TGG GTG TCC CTG<br>Pro Glu Pro Asp Gln Arg Arg Thr Leu Pro Pro Gly Trp Val Ser Leu<br>10 15 20    | 144 |
| GGC CGT GCG GAC CCT GAG GAA GAG CTG AGT CTC ACC TTT GCC CTG AGA<br>Gly Arg Ala Asp Pro Glu Glu Glu Leu Ser Leu Thr Phe Ala Leu Arg<br>25 30 35    | 192 |
| CAG CAG AAT GTG GAA AGA CTC TCG GAG CTG GTG CAG GCT GTG TCG GAT<br>Gln Gln Asn Val Glu Arg Leu Ser Glu Leu Val Gln Ala Val Ser Asp<br>40 45 50    | 240 |
| CCC AGC TCT CCT CAA TAC GGA AAA TAC CTG ACC CTA GAG AAT GTG GCT<br>Pro Ser Ser Pro Gln Tyr Gly Lys Tyr Leu Thr Leu Glu Asn Val Ala<br>55 60 65 70 | 288 |
| GAT CTG GTG AGG CCA TCC CCA CTG ACC CCG<br>Asp Leu Val Arg Pro Ser Pro Leu Thr Pro<br>75 80                                                       | 318 |

## (2) INFORMATION FOR SEQ ID NO: 53:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 329 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Normal prostate

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 69..140
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.9  
seq LCFLLLAVAMSFF/GS

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

```

AAGTTTCTGG AGCTGTTCCG AGTCCCGTGG AGTCTCCATC TGAGCCCTTT CCTAGTCCAG 60
GCATCCCG ATG TTG GTG GAT GGC CCA TCT GAG CGG CCA GCC CTG TGC TTC 110
 Met Leu Val Asp Gly Pro Ser Glu Arg Pro Ala Leu Cys Phe
 -20 -15

TTG CTG TTG GCT GTG GCA ATG TCT TTC TTC GGC TCA GCT CTA TCC ATA 158
Leu Leu Leu Ala Val Ala Met Ser Phe Phe Gly Ser Ala Leu Ser Ile
-10 -5 1 5

GAT GAA ACA CGG GCG CAT CTG TTG TTG AAA GAD AAG ATG ATG CGG CTG 206
Asp Glu Thr Arg Ala His Leu Leu Leu Lys Xaa Lys Met Met Arg Leu
 10 15 20

GGG GGG CGG CTG GTG CTG AAC ACC AAG GAG GAG CTG GCC AAT GAG AGG 254
Gly Gly Arg Leu Val Leu Asn Thr Lys Glu Glu Leu Ala Asn Glu Arg
 25 30 35

CTC ATG ACG CTC AAW ATC GCT GAG ATG AAG GAG GCC ATG AGG ACC CTG 302
Leu Met Thr Leu Xaa Ile Ala Glu Met Lys Glu Ala Met Arg Thr Leu
 40 45 50

ATA TTC CCA CCC AGC ATG CAC TTT TTC 329
Ile Phe Pro Pro Ser Met His Phe Phe
 55 60

```

## (2) INFORMATION FOR SEQ ID NO: 54:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 392 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Brain

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 9..59  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 6.8  
seq LVLVLVVAVTVRA/AL

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

```

AAGTTATC ATG GCG GCT CCC TTG GTC CTG GTG CTG GTG GTG GCT GTG ACA 50
 Met Ala Ala Pro Leu Val Leu Val Leu Val Val Ala Val Thr
 -15 -10 -5

GTG CGG GCG GCC TTG TTC CGC TCC AGT CTG GCC GAG TTC ATT TCC GAG 98
Val Arg Ala Ala Leu Phe Arg Ser Ser Leu Ala Glu Phe Ile Ser Glu
 1 5 10

CGG GTG GAG GTG GTG TCC CCA CTG AGC TCT TGG AAG AGA GTG GTT GAA 146
Arg Val Glu Val Val Ser Pro Leu Ser Ser Trp Lys Arg Val Val Glu
 15 20 25

GGC CTT TCA CTG TTG GAC TTG GGA GTA TCT CCG TAT TCT GGA GCA GTA 194
Gly Leu Ser Leu Leu Asp Leu Gly Val Ser Pro Tyr Ser Gly Ala Val
 30 35 40 45

TTT CAT GAA ACT CCA TTA ATA ATA TAC CTC TTT CAT TTC CTA ATT GAC 242
Phe His Glu Thr Pro Leu Ile Ile Tyr Leu Phe His Phe Leu Ile Asp
 50 55 60

TAT GCT GAA TTG GTG TTT ATG ATA ACT GAT GCA CTG ACT GCT ATT GCC 290
Tyr Ala Glu Leu Val Phe Met Ile Thr Asp Ala Leu Thr Ala Ile Ala
 65 70 75

CTG TAT TTT GCA ATC CAG GAC TTC AAT AAA GTT GTG TTT AAA AAG CAG 338
Leu Tyr Phe Ala Ile Gln Asp Phe Asn Lys Val Val Phe Lys Lys Gln
 80 85 90

AAA CTC CTC CTA GAA CTG GAC CAG TAT GCC CCA GAT GTG GCC GAA CTC 386
Lys Leu Leu Leu Glu Leu Asp Gln Tyr Ala Pro Asp Val Ala Glu Leu
 95 100 105

ATC CGG 392
Ile Arg
110

```

## (2) INFORMATION FOR SEQ ID NO: 55:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 418 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Substantia nigra

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: 23..328

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 6.7  
seq LXMTLMLPFKILS/DS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

|                                                                 |     |
|-----------------------------------------------------------------|-----|
| AGCTCATTG TAGGCTGAAC TA ATG ACT GCC GCC ATA AGA AGA CAG AGA GAA | 52  |
| Met Thr Ala Ile Arg Arg Gln Arg Glu                             |     |
| -100 -95                                                        |     |
| CTG AGT ATC CTC CCA AAG GTG ACA CTG GAA GCA ATG AAC ACC ACA GTG | 100 |
| Leu Ser Ile Leu Pro Lys Val Thr Leu Glu Ala Met Asn Thr Thr Val |     |
| -90 -85 -30                                                     |     |
| ATG CAA GGC TTC AAC AGA TCT GAG CGG TGC CCC AGA GAC ACT CGG ATA | 148 |
| Met Gln Gly Phe Asn Arg Ser Glu Arg Cys Pro Arg Asp Thr Arg Ile |     |
| -75 -70 -65                                                     |     |
| GTA CAG CTG GTA TTC CCA GCC CTC TAC ACA GTG GTT TTC TTG ACC GGC | 196 |
| Val Gln Leu Val Phe Pro Ala Leu Tyr Thr Val Val Phe Leu Thr Gly |     |
| -60 -55 -50 -45                                                 |     |
| ATC CTG CTG AAT ACT TTG GCT CTG TGG GTG TTT GTT CAC ATC CCC AGC | 244 |
| Ile Leu Leu Asn Thr Leu Ala Leu Trp Val Phe Val His Ile Pro Ser |     |
| -40 -35 -30                                                     |     |
| TCC TCC ACC TTC ATC ATC TAC CTC AAA AAC ACT TTG GTG GCC GAC TTG | 292 |
| Ser Ser Thr Phe Ile Ile Tyr Leu Lys Asn Thr Leu Val Ala Asp Leu |     |
| -25 -20 -15                                                     |     |
| ATN ATG ACA CTC ATG CTT CCT TTC AAA ATC CTC TCT GAC TCA CAC CTG | 340 |
| Xaa Met Thr Leu Met Leu Pro Phe Lys Ile Leu Ser Asp Ser His Leu |     |
| -10 -5 1                                                        |     |
| GCA CCC TGG CAG CTC AGA GCT TTT GTG TGT CGT TTT TCT TCG GTG ATA | 388 |
| Ala Pro Trp Gln Leu Arg Ala Phe Val Cys Arg Phe Ser Ser Val Ile |     |
| 5 10 15 20                                                      |     |
| TTT TAT GAG ACC ATG TAT GTG GGC GAG GGG                         | 418 |
| Phe Tyr Glu Thr Met Tyr Val Gly Glu Gly                         |     |
| 25 30                                                           |     |

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 379 base pairs

(B) TYPE: NUCLEIC ACID

(C) STRANDEDNESS: DOUBLE

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Spleen

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 203..340  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 6.3  
seq SIGVLTLSHLISG/LR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

```

ACTTTTTTCGG AGGGTGGTGA GCTAGTAAGT GTGGTTTTAG CTGTAGTAGC CAGATTGGGC 60
GCCCGGGAGT GGTGGGGGTG CCGGGTGGAA GGCTCTGGGC GGGGTCTCAG GACCCTCCTT 120
TTCTTGCGCG GGATCGGGCT TGTGGTGCCG CTCCCCGTAA TGTACGGAGG AAGAGGGAAA 180
GGGCTCTGGC CCCCTCGGCG TC ATG TCT TCG GTG CTG GCG GCT TCC CAT CCG 232
 Met Ser Ser Val Leu Ala Ala Ser His Pro
 -45 -40

CTG GTT CTA TCC TCA AAC GCC GGG ACA CCG GGA ATC TCG GAG AAG GAC 280
Leu Val Leu Ser Ser Asn Ala Gly Thr Pro Gly Ile Ser Glu Lys Asp
-35 -30 -25

AAC CGA GAT CCA GCT GGC TCC TCC ATC GGG GTG CTC ACA CTT TCT CAT 328
Asn Arg Asp Pro Ala Gly Ser Ser Ile Gly Val Leu Thr Leu Ser His
-20 -15 -10 -5

TTG ATT TCA GGT CTG CGG ACG CTG TAT ACC CTC CTC CAC TTC CCG CTG 376
Leu Ile Ser Gly Leu Arg Thr Leu Tyr Thr Leu Leu His Phe Pro Leu
 1 5 10

CGG 379
Arg

```

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 369 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Thyroid

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 55..204

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 6.3  
seq LIILGLVLFMVYGV/NV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

```

AGMGCAGGCC TGGTGGTGAG CAGGGACGGT GCACCGGACG GCGGGATCGA GCAA ATG 57
 Met
 -50

GGT CTG GCC ATG GAG CAC GGA GGG TCC TAC GCT CGG GCG GGG GGC AGC 105
Gly Leu Ala Met Glu His Gly Gly Ser Tyr Ala Arg Ala Gly Gly Ser
 -45 -40 -35

TCT CGG GGC TGC TGG TAT TAC CTG CGC TAC TTC TTC CTC TTC GTC TCC 153
Ser Arg Gly Cys Trp Tyr Tyr Leu Arg Tyr Phe Phe Leu Phe Val Ser
 -30 -25 -20

CTC ATC CAA TTC CTC ATC ATC CTG GGG CTC GTG CTC TTC ATG GTC TAT 201
Leu Ile Gln Phe Leu Ile Ile Leu Gly Leu Val Leu Phe Met Val Tyr
 -15 -10 -5

GGM AAC GTG CAC GTG AGC ACA GAG TCC AAC CTG CAG GCC ACC GAG CGC 249
Gly Asn Val His Val Ser Thr Glu Ser Asn Leu Gln Ala Thr Glu Arg
 1 5 10 15

CGA GCC GAG GGC CTA TAC AKY CAG CTC CTA GGG CTC ACG GCC TCC CAG 297
Arg Ala Glu Gly Leu Tyr Xaa Gln Leu Leu Gly Leu Thr Ala Ser Gln
 20 25 30

TCC AAC TTG ACC AAG GAG CTC AAC TTC ACC ACC CGC GCC AAG GAT GCC 345
Ser Asn Leu Thr Lys Glu Leu Asn Phe Thr Thr Arg Ala Lys Asp Ala
 35 40 45

ATC ATG CAG ATG TGG CTG AAT GCT 369
Ile Met Gln Met Trp Leu Asn Ala
 50 55

```

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 402 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 205..396
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.3  
seq SCLVSGWGLLANG/QR



## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

```

AAAAACGGCG AGGACTGCAG CCCGCACTCG CAGCCCTGGC AGGCGGCACT GGTCATGGAA 60
AACGAATTGT TCTGCTCGGG CGTCCTGGTG CATCCGCACT GGGTGCTGTC AGCCGCACAC 120
TGTTTCCAGA AGTGAGTKCA GAGCTCCTAC ACCATCGGGC TGGGCCTGCA CAGTCTTGAG 180
GCCGACCAAG AGCCAGGGAG CCAG ATG GTG GAG GCC AGC CTC TCC GTA CGG 231
 Met Val Glu Ala Ser Leu Ser Val Arg
 -60

CAC CCA GAG TAC AAC AGA CCC TTG CTC GCT AAC GAC CTC ATG CTC ATC 279
His Pro Glu Tyr Asn Arg Pro Leu Leu Ala Asn Asp Leu Met Leu Ile
-55 -50 -45 -40

AAG TTG GAC GAA TCC GTG TCC GAG TCT GAC ACC ATC CGG AGC ATC AGC 327
Lys Leu Asp Glu Ser Val Ser Glu Ser Asp Thr Ile Arg Ser Ile Ser
 -35 -30 -25

ATT GCT TCG CAG TGC CCT ACC GCG GGG AAC TCT TGC CTC GTT TCT GGC 375
Ile Ala Ser Gln Cys Pro Thr Ala Gly Asn Ser Cys Leu Val Ser Gly
 -20 -15 -10

TGG GGT CTG CTG GCG AAC GGC CAG CGG 402
Trp Gly Leu Leu Ala Asn Gly Gln Arg
 -5 1

```

## (2) INFORMATION FOR SEQ ID NO: 59:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 445 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Cancerous prostate

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 20..160
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.3  
seq VICCVLFLLFILG/YI

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

```

ACACTCCGGA GACTGAGCC ATG GGG GGA AAG CAG CGG GAC GAG GAT GAC GAG 52
 Met Gly Gly Lys Gln Arg Asp Glu Asp Asp Glu
 -45 -40

GCC TAC GGS AAG CCA GTC AAA TAC GAC CCC TCC TTT CGA GGC CCC ATC 100

```

|                                                                 |     |
|-----------------------------------------------------------------|-----|
| Ala Tyr Gly Lys Pro Val Lys Tyr Asp Pro Ser Phe Arg Gly Pro Ile |     |
| -35 -30 -25                                                     |     |
| AAG AAC AGA AGC TGC ACA GAT GTC ATC TGC TGC GTC CTC TTC CTG CTC | 148 |
| Lys Asn Arg Ser Cys Thr Asp Val Ile Cys Cys Val Leu Phe Leu Leu |     |
| -20 -15 -10 -5                                                  |     |
| TTC ATT CTA GGT TAC ATC GTG GTG GGG ATT GTG GCC TGG TTG TAT GGA | 196 |
| Phe Ile Leu Gly Tyr Ile Val Val Gly Ile Val Ala Trp Leu Tyr Gly |     |
| 1 5 10                                                          |     |
| GAC CCC CGG CAA GTC CTC TAC CCC AGG AAC TCT ACT GGG GCC TAC TGT | 244 |
| Asp Pro Arg Gln Val Leu Tyr Pro Arg Asn Ser Thr Gly Ala Tyr Cys |     |
| 15 20 25                                                        |     |
| GGC ATG GGG GAG AAC AAA GAT AAG CCG TAT CTC CTG TAC TTC AAC ATC | 292 |
| Gly Met Gly Glu Asn Lys Asp Lys Pro Tyr Leu Leu Tyr Phe Asn Ile |     |
| 30 35 40                                                        |     |
| TTC AGC TGC ATC CTG TCC AGC AAC ATC ATC TCA GTT GCT GAG AAC GGC | 340 |
| Phe Ser Cys Ile Leu Ser Ser Asn Ile Ile Ser Val Ala Glu Asn Gly |     |
| 45 50 55 60                                                     |     |
| CTA CAG TGC CCC ACA CCC CAG GTG TGT GTG TCC TCC TGC CCG GAG GAC | 388 |
| Leu Gln Cys Pro Thr Pro Gln Val Cys Val Ser Ser Cys Pro Glu Asp |     |
| 65 70 75                                                        |     |
| CCA TGG ACT NDB GRA AAA ACG AGT TCT CAC AGA CTG TTG GGG AAG TCT | 436 |
| Pro Trp Thr Xaa Xaa Lys Thr Ser Ser His Arg Leu Leu Gly Lys Ser |     |
| 80 85 90                                                        |     |
| TCT ATA CAA                                                     | 445 |
| Ser Ile Gln                                                     |     |
| 95                                                              |     |

## (2) INFORMATION FOR SEQ ID NO: 60:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 382 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Hypertrophic prostate

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 23..76
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.3  
seq VLLFLAWVCFLFY/AG

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

```

AACTTCCGGG TGCCATTGCA GG ATG CAG AAA GCC TCA GTG TTG CTC TTC CTG 52
 Met Gln Lys Ala Ser Val Leu Leu Phe Leu
 -15 -10

GCC TGG GTC TGC TTC CTC TTC TAC GCT GGC ATT GCC CTC TTC ACC AGT 100
Ala Trp Val Cys Phe Leu Phe Tyr Ala Gly Ile Ala Leu Phe Thr Ser
 -5 1 5

GGC TTC CTG CTC ACC CGT TTG GAR CTC ACC AAC CAT AGC AGC TGC CAA 148
Gly Phe Leu Leu Thr Arg Leu Glu Leu Thr Asn His Ser Ser Cys Gln
 10 15 20

GAG CCC CCA GGC CCT GGG TCC CTG CCA TGG GGG AGC CAA GGG AAA CCT 196
Glu Pro Pro Gly Pro Gly Ser Leu Pro Trp Gly Ser Gln Gly Lys Pro
 25 30 35 40

GGG GCC TGC TGG ATG GCT TCC CGA TTT TCG CGG GTT GTG TTG GTG CTG 244
Gly Ala Cys Trp Met Ala Ser Arg Phe Ser Arg Val Val Leu Val Leu
 45 50 55

ATA GAT GCT CTG CGA TTT GAC TTC GCC CAG CCC CAG CAT TCA CAC GTG 292
Ile Asp Ala Leu Arg Phe Asp Phe Ala Gln Pro Gln His Ser His Val
 60 65 70

CCT AGA GAG CCT CCT GTC TCC CTA CCC TTC CTG GGC AAA CTA AGC TCC 340
Pro Arg Glu Pro Pro Val Ser Leu Pro Phe Leu Gly Lys Leu Ser Ser
 75 80 85

TTG CAG AGG ATC CTG GAG ATT CAG CCC CAC CAT GCC CGG CTC 382
Leu Gln Arg Ile Leu Glu Ile Gln Pro His His Ala Arg Leu
 90 95 100

```

## (2) INFORMATION FOR SEQ ID NO: 61:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 402 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lung

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 133..375
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.7  
seq CWMMLLGSXGSFL/AP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

```

AAAAACGCGC GCSACGATTC GAGGTGCTCT GTGGCCGCGA GTGCATCTTC CACGAACCTA 60
ATTCATCTCT CCAGCAAAGG ACACATCTCT CCAGCAAAGG ACACCTCTCT CCAGCAAAGG 120

```

ACACCTGCAG AG ATG TCC CCA GTC CTT CAC TTC TAT GTT CGT CCC TCT GGC 171  
 Met Ser Pro Val Leu His Phe Tyr Val Arg Pro Ser Gly  
 -80 -75 -70

CAT GAG GGG GCA GCC TCT GGA CAC ACT CGG AGG AAA CTG CAA GGG AAA 219  
 His Glu Gly Ala Ala Ser Gly His Thr Arg Arg Lys Leu Gln Gly Lys  
 -65 -60 -55

CTG CCA GAG CTG CAG GGC GTC GAG ACT GAA CTG TGC TAC AAC GTG AAC 267  
 Leu Pro Glu Leu Gln Gly Val Glu Thr Glu Leu Cys Tyr Asn Val Asn  
 -50 -45 -40

TGG ACA GCT GAG GCC CTC CCC AGT GCT GAG GAG ACA AAG AAG CTG ATG 315  
 Trp Thr Ala Glu Ala Leu Pro Ser Ala Glu Glu Thr Lys Lys Leu Met  
 -35 -30 -25

TGG CTG TTT GGT TGC CCT TAC TGC TGG ATG ATG TTG CTC GGG AGT SCT 363  
 Trp Leu Phe Gly Cys Pro Tyr Cys Trp Met Met Leu Leu Gly Ser Xaa  
 -20 -15 -10 -5

GGC TCC TTC CTG GCT CCA ATG ACC TGC WGC TGG AGG TCG 402  
 Gly Ser Phe Leu Ala Pro Met Thr Cys Xaa Trp Arg Ser  
 1 5

## (2) INFORMATION FOR SEQ ID NO: 62:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 347 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Brain

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 114..221
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.6  
seq ILRLLGSLSNAYS/PR

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

GGAASYSGA CGCATGCGCC GTTCTCTGC ATGGTGTGCG TTCTCGTTCT AGCTGCGGCC 60

GCAGAGCTGT GGCGGTTTTT CTAATCCTGC GAATATGGGT AGTGCWTCGT TCC ATG 116  
 Met

GAC GTW ACG CCC CGG GAG TCT CTC AGT ATC TTG GTA GTG GCT GGG TCC 164  
 Asp Val Thr Pro Arg Glu Ser Leu Ser Ile Leu Val Val Ala Gly Ser  
 -35 -30 -25 -20

GGT GGG CAT ACC ACT GAG ATC CTG AGG CTG CTT GGG AGC TTG TCC AAT 212

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Gly | Gly | His | Thr | Thr | Glu | Ile | Leu | Arg | Leu | Gly | Ser | Leu | Ser | Asn |     |     |
|     |     |     |     | -15 |     |     |     |     | -10 |     |     |     |     | -5  |     |     |
| GCC | TAC | TCA | CCT | AGA | CAT | TAT | GTC | ATT | GCT | GAC | ACT | GAT | GAA | ATG | AGT | 260 |
| Ala | Tyr | Ser | Pro | Arg | His | Tyr | Val | Ile | Ala | Asp | Thr | Asp | Glu | Met | Ser |     |
|     |     |     | 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     |
| GCC | AAT | AAA | ATA | AAT | TCT | TTT | GAA | CTA | GAT | CGA | GCT | GAT | AGA | GAC | CCT | 308 |
| Ala | Asn | Lys | Ile | Asn | Ser | Phe | Glu | Leu | Asp | Arg | Ala | Asp | Arg | Asp | Pro |     |
|     | 15  |     |     |     |     | 20  |     |     |     |     | 25  |     |     |     |     |     |
| AGT | AAC | ATG | TAT | ACC | AAA | TAC | TAC | ATT | CAC | CGA | AAT | GGG |     |     |     | 347 |
| Ser | Asn | Met | Tyr | Thr | Lys | Tyr | Tyr | Ile | His | Arg | Asn | Gly |     |     |     |     |
| 30  |     |     |     |     | 35  |     |     |     |     | 40  |     |     |     |     |     |     |

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 451 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(D) DEVELOPMENTAL STAGE: Fetal  
(F) TISSUE TYPE: kidney

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 278..340  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 5.6  
seq LLRVLNLPHNSIG/CV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| ATACAAGCTC CACAGAGCCG CGGGAGGACG GTTGCCCTGGT ATTATTAGCA AGCAGCAAAT | 60  |
| ATGGCGGTGG CGCGCGTGGA CGCGGCTTTG CCTCCCGGAG AAGGATCAGT GGTCAATTGG  | 120 |
| TCAGGACARG GRMYWCCAGA AATTAGGTCC AAATTTACCC TGTGAAGCTG ATATTCACAC  | 180 |
| TTTGATTCTG GATAAAATC AGATTATTAA ATTGGAAAAT CTGGAGAAAT GCAAACGAWK   | 240 |
| AATACAGTTA TCAGTAGCTA ATAATCGGCT GGTTCCG ATG ATG GGT GTG GCC AAG   | 295 |
| Met Met Gly Val Ala Lys                                            |     |
| -20                                                                |     |
| CTG ACG TTG CTT CGT GTA TTA AAT TTG CCT CAT AAT AGC ATT GGC TGT    | 343 |
| Leu Thr Leu Leu Arg Val Leu Asn Leu Pro His Asn Ser Ile Gly Cys    |     |
| -15 -10 -5 1                                                       |     |
| GTG GAA GGG CTA AAG GAA CTA GTA CAT CTG GAA TGG CTG AAT TTG GCA    | 391 |
| Val Glu Gly Leu Lys Glu Leu Val His Leu Glu Trp Leu Asn Leu Ala    |     |
| 5 10 15                                                            |     |

GGA AAT AAT CTT AAG GCC ATG GAA CAG RTC AAT AGC TGC ACA GCT CTA 439  
 Gly Asn Asn Leu Lys Ala Met Glu Gln Xaa Asn Ser Cys Thr Ala Leu  
           20                          25                          30

CAG CAT CTC GAT 451  
 Gln His Leu Asp  
           35

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 333 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Brain

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 139..246
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.6  
seq ILRLGSLSNAYS/PR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

AACTTTGACA GCGGCTGGTC CCCGGAAGTT GK KYCGCATG CGCCGTTTCT CTGCATGGTG 60  
 TGC GTTCTCG TTCTAGCTGC GGCCGCAGAG CTGTGGCGGT TTTCCTAATC CTGCGAATAT 120  
 GGGGTAGTGC TTCGTTCC ATG GAC GTT ACG CCC CGG GAG TCT CTC AGT ATC 171  
                           Met Asp Val Thr Pro Arg Glu Ser Leu Ser Ile  
                           -35                          -30  
 TTG GTA GTG GCT GGG TCC GGT GGG CAT ACC ACT GAG ATC CTG AGG CTG 219  
 Leu Val Val Ala Gly Ser Gly Gly His Thr Thr Glu Ile Leu Arg Leu  
 -25                          -20                          -15                          -10  
 CTT GGG AGC TTG TCC AAT GCC TAC TCA CCT AGA CAT TAT GTC ATT GCT 267  
 Leu Gly Ser Leu Ser Asn Ala Tyr Ser Pro Arg His Tyr Val Ile Ala  
                           -5                          1                          5  
 GAC ACT GAT GAA ATG AGT GCC AAT AAA ATA AAT TCT TTT GAA CTA GAT 315  
 Asp Thr Asp Glu Met Ser Ala Asn Lys Ile Asn Ser Phe Glu Leu Asp  
           10                          15                          20  
 CGA GCT GAT AGA GAC CGG 333  
 Arg Ala Asp Arg Asp Arg  
           25

## (2) INFORMATION FOR SEQ ID NO: 65:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 175 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Colon

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 83..121
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.5  
seq MVLLTMIARVADG/LP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

```
AATAACTGTT GTCGCGGCGG AGGAAGTGAG GACGCGGCCA AGGGCCTTCC GGGCCAGTGT 60
TGGATCCCTG TAGTTTGTGA AG ATG GTG TTG CTA ACA ATG ATC GCC CGA GTG 112
 Met Val Leu Leu Thr Met Ile Ala Arg Val
 -10 -5

GCG GAC GGG CTC CCG CTG GCC GCC TCG ATG CAG GAG GAA GTG AGG ACG 160
Ala Asp Gly Leu Pro Leu Ala Ala Ser Met Gln Glu Glu Val Arg Thr
 1 5 10

GCG CCA AGG GCA TTG 175
Ala Pro Arg Ala Leu
15
```

## (2) INFORMATION FOR SEQ ID NO: 66:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 410 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Cancerous prostate

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 144..284
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.3  
seq GCGMFTFLSSVXA/AV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

```

ACACAAATCA CATTAGCTTT GCCCGAAGTT TTTCCCCACA CTCTTCTTTA GCATGCTATT 60
ATGGGGAAAG TGACCACTCC TGGGAGCGGG GGTGGTCGGG GCGGTTTGGT GCGGGGGAAG 120
CGGCTGTAAC TTCTAMGKKR ACC ATG GTA CCT GTT GAA AAC ACC GAG GGC CCC 173
 Met Val Pro Val Glu Asn Thr Glu Gly Pro
 -45 -40

AGT CTG CTG AAC CAG AAG GGG ACA GCC GTG GAG ACG GAG GGC AKC GGC 221
Ser Leu Leu Asn Gln Lys Gly Thr Ala Val Glu Thr Glu Gly Xaa Gly
 -35 -30 -25

AGC CGG CAT CCT CCC TGG GCG AGA GGC TGC GGC ATG TTT ACC TTC CTG 269
Ser Arg His Pro Pro Trp Ala Arg Gly Cys Gly Met Phe Thr Phe Leu
 -20 -15 -10

TCA TCT GTC ANT GCT GCT GTC AGT GGC CTC CTG GTG GGT TAT GAA CTT 317
Ser Ser Val Xaa Ala Ala Val Ser Gly Leu Leu Val Gly Tyr Glu Leu
 -5 1 5 10

GGG ATC ATC TCT GGG GCT CTT CTT CAG ATC AAA ACC TTA TTA GCC NTG 365
Gly Ile Ile Ser Gly Ala Leu Leu Gln Ile Lys Thr Leu Leu Ala Xaa
 15 20 25

AGC TGC CAT GAG CAG GAA ATG GTT GTG AGC TCC CTC GTC ATT GGA 410
Ser Cys His Glu Gln Glu Met Val Val Ser Ser Leu Val Ile Gly
 30 35 40

```

## (2) INFORMATION FOR SEQ ID NO: 67:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 377 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Cancerous prostate

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 237..308
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.2  
seq LLFPVGRSWSCFA/QT

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

```

ACCTGTCTTG AGGTCTAATG GCGGACGCCA GSTATGTGGA GTTGGTGGTG GCTTAAGTTT 60
TGAAGGGAGG TAGCATCCGT TGGATATCCA CACCATCCTT CTCGCTGCAG GCTTTCTTGG 120

```



ACTCCGTACT GTTGGTGTAA CCAAGGCCTG GAGGTCTGGG TGGCTCAGGT TTCCTGCAGC 180  
 CATGTTTCTG TACAACCTAA CCTTGCAGAG AGQCACTGGC ATCAGCTTTG CCATTC ATG 239  
 Met  
 GAA ACT TTT CTG GAA CCA AAC AAC AAG AAA TTG TTG TTT CCC GTG GGA 287  
 Glu Thr Phe Leu Glu Pro Asn Asn Lys Lys Leu Leu Phe Pro Val Gly  
 -20 -15 -10  
 AGA TCT TGG AGC TGC TTC GCC CAG ACC CBN TCA CTG GCA AAG TAC ATA 335  
 Arg Ser Trp Ser Cys Phe Ala Gln Thr Xaa Ser Leu Ala Lys Tyr Ile  
 -5 1 5  
 CCC TAC TCA CTG TGG AAG TAT TCG GTG TTA TCC GGT CAC TCA 377  
 Pro Tyr Ser Leu Trp Lys Tyr Ser Val Leu Ser Gly His Ser  
 10 15 20

## (2) INFORMATION FOR SEQ ID NO: 68:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 360 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Normal prostate

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 31..75
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.1  
seq FLWGLALPLFFFC/WE

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

AGTTCTGTGG AGCAGCGGTG GCCGGCTAGG ATG GGC TTT CTC TGG GGT CTG GCT 54  
 Met Gly Phe Leu Trp Gly Leu Ala  
 -15 -10  
 CTG CCC CTT TTC TTC TTC TGC TGG GAG GTT GGG GTC TCT GGG AGC TCT 102  
 Leu Pro Leu Phe Phe Phe Cys Trp Glu Val Gly Val Ser Gly Ser Ser  
 -5 1 5  
 GCA GGC CCC AGC ACC CGC AGA GCA GAC ACT GCG ATG ACA ACG GAC GAC 150  
 Ala Gly Pro Ser Thr Arg Arg Ala Asp Thr Ala Met Thr Thr Asp Asp  
 10 15 20 25  
 ACA GAA GTG CCC GCT ATG ACT CTA GCA CCG GGC CAC GCC GCT CTG GAA 198  
 Thr Glu Val Pro Ala Met Thr Leu Ala Pro Gly His Ala Ala Leu Glu  
 30 35 40  
 ACT CAA ACA CTG AGC GCT GAG ACC TCT TCT AGG GCC TCA ACC CCA GCC 246  
 Thr Gln Thr Leu Ser Ala Glu Thr Ser Ser Arg Ala Ser Thr Pro Ala

|                                                                 |    |    |     |
|-----------------------------------------------------------------|----|----|-----|
| 45                                                              | 50 | 55 |     |
| GGC CCC GTT CCA GAA GCA GAG ACC AGG GGA GCC AAG AGA ATT TCC CCT |    |    | 294 |
| Gly Pro Val Pro Glu Ala Glu Thr Arg Gly Ala Lys Arg Ile Ser Pro |    |    |     |
| 60                                                              | 65 | 70 |     |
| GCA AGA GAG ACC AGG AGT TTC ACA AAA ACR KHK CCC AAC TTC ATG GTG |    |    | 342 |
| Ala Arg Glu Thr Arg Ser Phe Thr Lys Thr Xaa Pro Asn Phe Met Val |    |    |     |
| 75                                                              | 80 | 85 |     |
| CTG AGN DAN ANC GTC ACG                                         |    |    | 360 |
| Leu Xaa Xaa Xaa Val Thr                                         |    |    |     |
| 90                                                              | 95 |    |     |

## (2) INFORMATION FOR SEQ ID NO: 69:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 339 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Spleen

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 106..168
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.9  
seq WLLSDILGQGATA/NV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

|                                                                          |     |
|--------------------------------------------------------------------------|-----|
| AAAGCCGGAA GTGTCCTGAG TCTCGAGGAG GCCGCGGGAG CCCGCCGCGC GTGGCGCGGC        | 60  |
| GGAGACCCGG CTGGTATAAC AAGAGGATTG CCTGATCCAG CCAAG ATG CAG AGC ACT        | 117 |
| Met Gln Ser Thr                                                          |     |
| -20                                                                      |     |
| TCT AAT CAT CTG TGG CTT TTA TCT GAT ATT TTA GGC CAA GGA GCT ACT          | 165 |
| Ser Asn His Leu Trp Leu Leu Ser Asp Ile Leu Gly Gln Gly Ala Thr          |     |
| -15                      -10                      -5                     |     |
| GCA AAT GTC TTT CGT GGA AGA CAT AAG AAA ACT GGT GAT TTA TTT GCT          | 213 |
| Ala Asn Val Phe Arg Gly Arg His Lys Lys Thr Gly Asp Leu Phe Ala          |     |
| 1                      5                      10                      15 |     |
| ATC AAA GTA TTT AAT AAC ATA AGC TTC CTT CGT CCA GTG GAT GTT CAA          | 261 |
| Ile Lys Val Phe Asn Asn Ile Ser Phe Leu Arg Pro Val Asp Val Gln          |     |
| 20                      25                      30                       |     |
| ATG AGA GAA TTT GAA GTG TTG AAA AAA CTC AAT CAC AAA AAT ATT GTC          | 309 |
| Met Arg Glu Phe Glu Val Leu Lys Lys Leu Asn His Lys Asn Ile Val          |     |
| 35                      40                      45                       |     |

AAA TTA TTT GCT ATT GAA GAA GAG ACA GGG  
 Lys Leu Phe Ala Ile Glu Glu Glu Thr Gly  
           50                              55

339

## (2) INFORMATION FOR SEQ ID NO: 70:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 236 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 120..167
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.9  
seq ICAGSVLPPYSNC/QM

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

AAACCCTGGT GTTCCTGACA CAAACTTCAG GAAAGGATTT TGCACTTGTG CAGACCGGGC 60  
 GAGCAGAGTA AGAAGCAGGT ACGTGGGTTT TTCCAAGTTC TGTGTTTCAG TCCTGTTGG 119  
 ATG GTT GAG ATC TGT GCA GGG TCT GTG CTT CCG CCT TAT TCA AAC TGT 167  
 Met Val Glu Ile Cys Ala Gly Ser Val Leu Pro Pro Tyr Ser Asn Cys  
       -15                              -10                              -5  
 CAG ATG CCA GAA CCT TCG ATC TTT ACT TTG ATA CAT TTC CAC ACT TAT 215  
 Gln Met Pro Glu Pro Ser Ile Phe Thr Leu Ile His Phe His Thr Tyr  
       1                              5                              10                              15  
 TAC TGC CTC ACA ACC CCA CAG 236  
 Tyr Cys Leu Thr Thr Pro Gln  
                               20

## (2) INFORMATION FOR SEQ ID NO: 71:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 255 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Brain

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 37..165  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 4.7  
 seq LLAFGTSCSVVXY/XP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

```

AGCGTCTCTT GTTTGTGCGG CTGACCAGTT GGCGAC ATG GTG GCA CCC GTG CTG 54
 Met Val Ala Pro Val Leu
 -40

GAG ACT TCT CAC GTG TTT TGC TGC CCA AAC CGG GTG CGG GGM GTC CTG 102
Glu Thr Ser His Val Phe Cys Cys Pro Asn Arg Val Arg Gly Val Leu
-35 -30 -25

AAC TGG WGC TCT GGG CCC AGA GGA CTT CTG GCC TTT GGC ACG TCC TGC 150
Asn Trp Xaa Ser Gly Pro Arg Gly Leu Leu Ala Phe Gly Thr Ser Cys
-20 -15 -10

TCC GTG GTG CKC TAT GRC CCC CTG AWM AGG GTT GTT GTT ACC ARC TTG 198
Ser Val Val Xaa Tyr Xaa Pro Leu Xaa Arg Val Val Val Thr Xaa Leu
-5 1 5 10

MAT GGT CAC ACC GCC CGA GTC AAT TGC ATA CAG TGG ATT KGT AAA CAG 246
Xaa Gly His Thr Ala Arg Val Asn Cys Ile Gln Trp Ile Xaa Lys Gln
15 20 25

GRA GGC ATG 255
Xaa Gly Met
30

```

## (2) INFORMATION FOR SEQ ID NO: 72:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 425 base pairs  
 (B) TYPE: NUCLEIC ACID  
 (C) STRANDEDNESS: DOUBLE  
 (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
 (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 75..284  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 4.7  
 seq QLLLATLQEATT/QE

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

```

AAGTGAGACC GCGCGGCAAC AGCTTGCGGC TCGGGGAGC TCCCGTGGGC GCTCCGCTGG 60
CTGTGCAGGC GGCC ATG GAT TCC TTG CGG AAA ATG CTG ATC TCA GTC GCA 110
 Met Asp Ser Leu Arg Lys Met Leu Ile Ser Val Ala
 -70 -65 -60

ATG CTG GGC GCA RGG GCT GGC GTG GGC TAC GCG CTC CTC GTT ATC GTG 158
Met Leu Gly Ala Xaa Ala Gly Val Gly Tyr Ala Leu Leu Val Ile Val
 -55 -50 -45

ACC CCG GGA GAG CGG CGG AAG CAG GAA ATG CTA AAG GAG ATG CCA CTG 206
Thr Pro Gly Glu Arg Arg Lys Gln Glu Met Leu Lys Glu Met Pro Leu
 -40 -35 -30

CAG GAC CCA AGG AGC AGG GAG GAG GCG GCC AGG ACC CAG CAG CTA TTG 254
Gln Asp Pro Arg Ser Arg Glu Glu Ala Ala Arg Thr Gln Gln Leu Leu
 -25 -20 -15

CTG GCC ACT CTG CAG GAG GCA GCG ACC ACG CAG GAG AAC GTG GCC TGG 302
Leu Ala Thr Leu Gln Glu Ala Ala Thr Thr Gln Glu Asn Val Ala Trp
 -10 -5 1 5

AGG AAG AAC TGG ATG GTT GGC GGC GAA GGC GGC GCC ACG GGA NNT CAC 350
Arg Lys Asn Trp Met Val Gly Gly Glu Gly Gly Ala Thr Gly Xaa His
 10 15 20

CGT GAG ACC GGA CTT GCV TCC GTG GGC GCC GGA CCT TGG CTT GGG CGC 398
Arg Glu Thr Gly Leu Ala Ser Val Gly Ala Gly Pro Trp Leu Gly Arg
 25 30 35

AGG AAT CCG AGG CAG CTT TCT CCT TCG 425
Arg Asn Pro Arg Gln Leu Ser Pro Ser
 40 45

```

## (2) INFORMATION FOR SEQ ID NO: 73:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 380 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (E) TISSUE TYPE: Hypertrophic prostate

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 108..185
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.6  
seq LLPFGMLCASSTT/KC

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

```

AACTTTCACT TTCGAGAGTG CCGTCTATTT GCCACACACT TCCCTGATGA AATGTCTGGA 60

TTTGGAATAA AGAAAAAAGG AAAGGCTAGC AGTCATCCAA CAGAATC ATG AGA CAG 116
 Met Arg Gln
 -25

ACT TTG CCT TGT ATC TAC TTT TGG GGG GGC CTT TTG CCC TTT GGG ATG 164
Thr Leu Pro Cys Ile Tyr Phe Trp Gly Gly Leu Leu Pro Phe Gly Met
 -20 -15 -10

CTG TGT GCA TCC TCC ACC ACC AAG TGC ACT GTT AGC CAT GAA GTT GCT 212
Leu Cys Ala Ser Ser Thr Thr Lys Cys Thr Val Ser His Glu Val Ala
 -5 1 5

GAC TGC AGC CAC CTG AAG TTG ACT CAG GTA CCC GAT GAT CTA CCC ACA 260
Asp Cys Ser His Leu Lys Leu Thr Gln Val Pro Asp Asp Leu Pro Thr
 10 15 20 25

AAC ATA ACA GTG TTG AAC CTT ACC CAT AAT CAA CTC AGA AGA TTA CCA 308
Asn Ile Thr Val Leu Asn Leu Thr His Asn Gln Leu Arg Arg Leu Pro
 30 35 40

GCC GCC AAC TTC ACA AGG TAT AGC CAG CTA ACT AGC TTG GAT GTA GGA 356
Ala Ala Asn Phe Thr Arg Tyr Ser Gln Leu Thr Ser Leu Asp Val Gly
 45 50 55

TTT AAC ACC ATC TCA AAA CTG GAG 380
Phe Asn Thr Ile Ser Lys Leu Glu
 60 65

```

## (2) INFORMATION FOR SEQ ID NO: 74:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 406 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Cancerous prostate

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 5..334
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.6  
seq HTXGLLGFRXQG/SI

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

```

AACT ATG GCC GAT GAT CTG GAG CAG CAG TCT CAA GGC TGG CTG AGT AGC 49
Met Ala Asp Asp Leu Glu Gln Gln Ser Gln Gly Trp Leu Ser Ser
-110 -105 -100

TGG CTG CCC ACC TGG CGC CCC ACT TCC ATG TCT CAG CTG AAG AAT GTG 97

```

Trp Leu Pro Thr Trp Arg Pro Thr Ser Met Ser Gln Leu Lys Asn Val  
 -95 -90 -85 -80

GAA GCC AGG ATC CTC CAG TGT CTC CAG AAT AAG TTC CTG GCC AGA TAT 145  
 Glu Ala Arg Ile Leu Gln Cys Leu Gln Asn Lys Phe Leu Ala Arg Tyr  
 -75 -70 -65

GTA TCC CTC CCA AAC CAG AAT AAG ATC TGG ACG GTG ACT GTG AGC CCC 193  
 Val Ser Leu Pro Asn Gln Asn Lys Ile Trp Thr Val Thr Val Ser Pro  
 -60 -55 -50

GAG CAA AAC GAC CGC ACC CCC TTG GTG ATG GTG CAT GGT TTT GGG GGC 241  
 Glu Gln Asn Asp Arg Thr Pro Leu Val Met Val His Gly Phe Gly Gly  
 -45 -40 -35

GGC GTG GGT CTC TGG ATC CTC AAC ATG GAC TCA CTG ART GCC CGC CGC 289  
 Gly Val Gly Leu Trp Ile Leu Asn Met Asp Ser Leu Xaa Ala Arg Arg  
 -30 -25 -20

ACA CTG CAC ACC TTH GGT CTG CTT GGC TTC GGG CGA AST CAA GGC AGC 337  
 Thr Leu His Thr Xaa Gly Leu Leu Gly Phe Gly Arg Xaa Gln Gly Ser  
 -15 -10 -5 1

ATT CCC AAG GGA CCG GAG GGG CTK RAG GAT GAG TTT GTG AMA TCR ATA 385  
 Ile Pro Lys Gly Pro Glu Gly Leu Xaa Asp Glu Phe Val Xaa Ser Ile  
 5 10 15

GRR ACA TGG CGG GAG ACA TGG 406  
 Xaa Thr Trp Arg Glu Thr Trp  
 20

## (2) INFORMATION FOR SEQ ID NO: 75:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 291 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (E) TISSUE TYPE: Large intestine

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 94..165
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.5  
seq PLSMILLSDKIQS/SK

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

ATCATACGAT CTACTTTTTT TAATGCCGTT GAAACAGAGT TAATTCCTT TAGCACACAA 60

GTCTTAGAGA CAAAGAAAAA AAAGGTCTGC AAC ATG AAA GTC ACA GGC ATC ACA 114  
 Met Lys Val Thr Gly Ile Thr

-20

ATC CTC TTT TGG CCC CTC TCC ATG ATA TTA TTA TCA GAC AAA ATC CAG 162  
 Ile Leu Phe Trp Pro Leu Ser Met Ile Leu Leu Ser Asp Lys Ile Gln  
 -15 -10 -5

TCT TCT AAA AGA GAA GTC CAA TGT AAT TTT ACT GAA AAA AAT TAT ACC 210  
 Ser Ser Lys Arg Glu Val Gln Cys Asn Phe Thr Glu Lys Asn Tyr Thr  
 1 5 10 15

TTG ATT CCA GCA GAT ATC AAG AAA GAT GTT ACT ATA CTT GAT CTC AGT 258  
 Leu Ile Pro Ala Asp Ile Lys Lys Asp Val Thr Ile Leu Asp Leu Ser  
 20 25 30

TAT AAC CAR VDB ACT CTT AAT GGC ACA GAC ACG 291  
 Tyr Asn Gln Xaa Thr Leu Asn Gly Thr Asp Thr  
 35 40

## (2) INFORMATION FOR SEQ ID NO: 76:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 327 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (E) TISSUE TYPE: Brain

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 7..294
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.4  
seq HLSWSSSAYQAWA/QE

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

AGCATC ATG GCG GCT GGC CGG GCC CAG GTC CCT TCC TCC GAA CAA GCC 48  
 Met Ala Ala Gly Arg Ala Gln Val Pro Ser Ser Glu Gln Ala  
 -95 -90 -85

TGG CTT GAG GAT GCT CAG GTC TTC ATC CAA AAG ACC CTG TGT CCA GCT 96  
 Trp Leu Glu Asp Ala Gln Val Phe Ile Gln Lys Thr Leu Cys Pro Ala  
 -80 -75 -70

GTC AAG GAG COT AAT GTC CAG TTG ACT CCA TTG GTA ATT GAT TGT GTG 144  
 Val Lys Glu Pro Asn Val Gln Leu Thr Pro Leu Val Ile Asp Cys Val  
 -65 -60 -55

AAG ACT GTC TGS TTG TCC CAG GGA AGG AAC CAA GGT TCT ACA CTG CCC 192  
 Lys Thr Val Trp Leu Ser Gln Gly Arg Asn Gln Gly Ser Thr Leu Pro  
 -50 -45 -40 -35

CTC AGC CAT ACC TTC GTC TCA GTA CAG GAC CTC AAG ACT CAC CAG CGT 240



Leu Ser Tyr Ser Phe Val Ser Val Gln Asp Leu Lys Thr His Gln Arg  
                   -30                  -25                  -20

CTC CCA TGC TGC AGC CAC CTG TCG TGG AGC AGT AGT GCA TAC CAG GCC 288  
 Leu Pro Cys Cys Ser His Leu Ser Trp Ser Ser Ser Ala Tyr Gln Ala  
                   -15                  -10                  -5

TGG GCC CAA GAG GCT GGA CCA AAT GGG AAC CCC CCT GGG 327  
 Trp Ala Gln Glu Ala Gly Pro Asn Gly Asn Pro Pro Gly  
           1                          5                          10

## (2) INFORMATION FOR SEQ ID NO: 77:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 311 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Normal prostate

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 186..227
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4  
seq STCCWCTPGGAST/ID

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

AACTTCCGCT GGTGGCCTAG AGCGGGGCCC GGTATGGAGG TGGGCTAGAG GCCGACGCCA 60  
 GCCAGAGAGC GAAATGTTCT TTTGGGGCCA GAGTCTGGGC ATATATGAAT GCAAATCCGT 120  
 GTTGTTCAC AACTAAGCCC AGCTGAGACG ATCACTTTC TGTAGGCCAT TTGTCCAGGT 180  
 ATAGA ATG AGC ACA TGT TGT TGG TGT ACG CCA GGT GGT GCT TCC ACC ATT 230  
   Met Ser Thr Cys Cys Trp Cys Thr Pro Gly Gly Ala Ser Thr Ile  
                   -10                  -5                  1

GAC TTC CTA AAG CGC TAT GCT TCC AAC ACT CCG TCC GGT GAA TTT CAA 278  
 Asp Phe Leu Lys Arg Tyr Ala Ser Asn Thr Pro Ser Gly Glu Phe Gln  
           5                          10                          15

ACA GCG GAG GAA GAC CTC TGC TAC TGC TTG GGG 311  
 Thr Ala Asp Glu Asp Leu Cys Tyr Cys Leu Gly  
           20                          25

## (2) INFORMATION FOR SEQ ID NO: 78:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 297 base pairs

- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 139..246
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.9  
seq VVEILPYLPCLTA/RD

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

```
ACTCCTCGCT GCGGGAAGGG TCCTGGGNCC CGGGCGGCGG TCGCCAGGTC TCAGGGCCGG 60
GGGTACCCGA GTCTCGTTTC CTCTCAGTCC ATCCACCCTT CATGGGGCCA GAGCCCTCTC 120
TCCAGAATCT GAGCAGCA ATG CCG TTT GCT GAA GAC AAG ACC TAT AAG TAT 171
 Met Pro Phe Ala Glu Asp Lys Thr Tyr Lys Tyr
 -35 -30
ATC TGC CGC AAT TTC AGC AAT TTT TGC DAT GTG GAT GTT GTA GAG ATT 219
Ile Cys Arg Asn Phe Ser Asn Phe Cys Xaa Val Asp Val Val Glu Ile
-25 -20 -15 -10
CTG CCT TAC CTG CCC TGC CTC ACA GCA AGA GAC CAG GAT CGA CTG CGG 267
Leu Pro Tyr Leu Pro Cys Leu Thr Ala Arg Asp Gln Asp Arg Leu Arg
 -5 1 5
GCC ACC TGC ACA CTC TCA GGG AAC CGG GCG 297
Ala Thr Cys Thr Leu Ser Gly Asn Arg Ala
 10 15
```

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 463 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 113..433
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.9

seq IVLVLLGRYTEE/EQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

```

AAAAAAGCAA AAGCAACAGC TCAAGCAGCC TCCTTGAGAGA AAACCTGAAA ATTCAACTTG 60
TTCAAGAGAA GGTCTTGTAC GTGCCTAAGT TCTAGAGCCT CCTGACGTGA GC ATG GCT 118
 Met Ala

GAG AGT GAG GAC CGC TCC CTG AGG ATC GTT CTG GTA GGG AAA ACT GGA 166
Glu Ser Glu Asp Arg Ser Leu Arg Ile Val Leu Val Gly Lys Thr Gly
-105 -100 -95 -90

AGT GGG AAA AGT GCA ACA GCG AAC ACC ATC CTT GGA GAG GAA ATC TTT 214
Ser Gly Lys Ser Ala Thr Ala Asn Thr Ile Leu Gly Glu Glu Ile Phe
 -85 -80 -75

GAT TCT AGA ATT GCT GCC CAA GCT GTT ACC AAG AAC TGT CAA AAA GCA 262
Asp Ser Arg Ile Ala Ala Gln Ala Val Thr Lys Asn Cys Gln Lys Ala
 -70 -65 -60

TCC CGG GAA TGG CAG GGG AGA GAC CTT CTT GTT GTG GAC ACT CCA GGG 310
Ser Arg Glu Trp Gln Gly Arg Asp Leu Leu Val Val Asp Thr Pro Gly
 -55 -50 -45

CTC TTT GAC ACC AAG GAG AGC CTG GAB ACC ACC TGC AAG GAA ATC RGC 358
Leu Phe Asp Thr Lys Glu Ser Leu Xaa Thr Thr Cys Lys Glu Ile Xaa
 -40 -35 -30

CGC TGC ATC ATC TCC TCC TGC CCA GGG CCC CAT GCT ATT GTC CTA GTT 406
Arg Cys Ile Ile Ser Ser Cys Pro Gly Pro His Ala Ile Val Leu Val
-25 -20 -15 -10

CTG CTG CTG GGC CGC TAC ACA GAG GAG GAG CAG AAA ACC GTT GCA TTG 454
Leu Leu Leu Gly Arg Tyr Thr Glu Glu Glu Gln Lys Thr Val Ala Leu
 -5 1 5

ATC ARG CTG 463
Ile Xaa Leu
10

```

(2) INFORMATION FOR SEQ ID NO: 80:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 369 base pairs
  - (B) TYPE: NUCLEIC ACID
  - (C) STRANDEDNESS: DOUBLE
  - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Homo Sapiens
  - (F) TISSUE TYPE: Lung
- (ix) FEATURE:
  - (A) NAME/KEY: sig\_peptide
  - (B) LOCATION: 73..219

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 3.8  
seq LLXCVGNFFGSTQ/DA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

```

AATTTTTC GGGGAACGCG GATTCGCATT CCCAATTTTA GGTGGCAGTC GCAACCCATA 60
CTATTCGGAC AG ATG GCA CAG AAA CCG CTG CGC CTC TTG GCT TGT GGA GAT 111
 Met Ala Gln Lys Pro Leu Arg Leu Leu Ala Cys Gly Asp
 -45 -40

GTT GAA GGA AAG TTT GAT ATT TTA TTC AAT AGA GTT CAA GCA ATT CAG 159
Val Glu Gly Lys Phe Asp Ile Leu Phe Asn Arg Val Gln Ala Ile Gln
 -35 -30 -25

AAG ARR AGT GGA AAC TTT GAT CTG CTG TKG TGT GTA GGA AAT TTC TTT 207
Lys Xaa Ser Gly Asn Phe Asp Leu Leu Xaa Cys Val Gly Asn Phe Phe
 -20 -15 -10 -5

GGC TCC ACC CAA GAT GCT GAA TGG GAG GAG TAT AAG ACT GGC ATC AAG 255
Gly Ser Thr Gln Asp Ala Glu Trp Glu Glu Tyr Lys Thr Gly Ile Lys
 1 5 10

AAA GCT CCT ATT CAG ACA TAT GTG CTT GGT GCT AAT AAC CAG GAA ACA 303
Lys Ala Pro Ile Gln Thr Tyr Val Leu Gly Ala Asn Asn Gln Glu Thr
 15 20 25

GTA AAA TAT TTC CAG GAT GCT GAT GGA TGT GAA TTA GCT GAA AAC ATT 351
Val Lys Tyr Phe Gln Asp Ala Asp Gly Cys Glu Leu Ala Glu Asn Ile
 30 35 40

ACT TAT CTG GGG CGA GGG 369
Thr Tyr Leu Gly Arg Gly
 45 50

```

(2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 383 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 57..212
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.8  
seq RPVLLHLHQTAHA/DE

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

ACGGTCAAGC TAAGGCGAAG AGTGGGTGGC TGAAGCCATA CTATTTTATA GAATTA ATG 59  
Met

GAA AGC AGA AAA GAC ATC ACA AAC CAA GAA GAA CTT TGG AAA ATG AAG 107  
Glu Ser Arg Lys Asp Ile Thr Asn Gln Glu Glu Leu Trp Lys Met Lys  
-50 -45 -40

CCT AGG AGA AAT TTA GAA GAA GAC GAT TAT TTG CAT AAG GAC ACG GGA 155  
Pro Arg Arg Asn Leu Glu Glu Asp Asp Tyr Leu His Lys Asp Thr Gly  
-35 -30 -25 -20

GAG ACC AGC ATG CTA AAA AGA CCT GTG CTT TTG CAT TTG CAC CAA ACA 203  
Glu Thr Ser Met Leu Lys Arg Pro Val Leu Leu His Leu His Gln Thr  
-15 -10 -5

GCC CAT GCT GAT GAA TTT GAC TGC CCT TCA GAA CTT CAG CAC ACA CAG 251  
Ala His Ala Asp Glu Phe Asp Cys Pro Ser Glu Leu Gln His Thr Gln  
1 5 10

CAA CTC TTT CCA CAG TGG CAC TTG CCA ATT AAA ATA GCT GCT ATT ATA 299  
Gln Leu Phe Pro Gln Trp His Leu Pro Ile Lys Ile Ala Ala Ile Ile  
15 20 25

GCA WCT CTG ACT TTT CTT TAC ACT CTT CTG AGG GAA GTA ANT CAC CCT 347  
Ala Xaa Leu Thr Phe Leu Tyr Thr Leu Leu Arg Glu Val Xaa His Pro  
30 35 40 45

TTA GCA ACT TCC CAT CAA CAA TAT TTT TAT AAA ATT 383  
Leu Ala Thr Ser His Gln Gln Tyr Phe Tyr Lys Ile  
50 55

## (2) INFORMATION FOR SEQ ID NO: 82:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 277 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Testis

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 80..235
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.8  
seq RPVLLHLHQTAHA/DE

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

AGCAGCAGCG GCAGCCGAGA CTCACGGTCA AGCTAAGGCG AAGAGTGGGT GGCTGAAGCC 60

(2) INFORMATION FOR SEQ ID NO: 83:

(A) LENGTH: 358 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Colon

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 92..199  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 3.7  
seq STLASVPPAATFG/AD

|            |            |            |      |      |        |            |            |     |     |     |     |     |     |     |     |     |
|------------|------------|------------|------|------|--------|------------|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| AAGATACCTC | AGCGCTACCT | GCGGGA     | ACTG | GATT | TCTCTC | CCGCCTGCCG | GCCTGCCTGC | 60  |     |     |     |     |     |     |     |     |
| CACAGCCGGA | CTCCGCCACT | COGGTAGCCT | C    | ATG  | GCT    | GCA        | ACC        | TGT | GAG | ATT | 112 |     |     |     |     |     |
|            |            |            |      | Met  | Ala    | Ala        | Thr        | Cys | Glu | Ile |     |     |     |     |     |     |
|            |            |            |      | -35  |        |            |            |     |     | -30 |     |     |     |     |     |     |
| AGC        | AAC        | ATT        | TTT  | AGC  | AAC    | TAC        | TTC        | AGT | GCG | ATG | TAC | AGC | TCG | GAG | GAC | 160 |
| Ser        | Asn        | Ile        | Phe  | Ser  | Asn    | Tyr        | Phe        | Ser | Ala | Met | Tyr | Ser | Ser | Glu | Asp |     |
|            |            |            |      | -25  |        |            |            |     | -20 |     |     |     |     | -15 |     |     |
| TCC        | ACC        | CTG        | GCC  | TCT  | GTT    | CCC        | CCT        | GCT | GCC | ACC | TTT | GGG | GCC | GAT | GAC | 208 |
| Ser        | Thr        | Leu        | Ala  | Ser  | Val    | Pro        | Pro        | Ala | Ala | Thr | Phe | Gly | Ala | Asp | Asp |     |
|            |            |            | -10  |      |        |            |            | -5  |     |     |     |     | 1   |     |     |     |
| TTG        | GTA        | CTG        | ACC  | CTG  | AGC    | AAC        | CCC        | CAG | ATG | TCA | TTG | GAG | GGT | ACA | GAG | 256 |

```

Leu Val Leu Thr Leu Ser Asn Pro Gln Met Ser Leu Glu Gly Thr Glu
 5 10 15
AAG GCC AGC TGG TTG GGG GAA CAG CCC CAG THC TGG TCG AAG ACG CAG 304
Lys Ala Ser Trp Leu Gly Glu Gln Pro Gln Xaa Trp Ser Lys Thr Gln
 20 25 30 35
GTT CTG GAC TGG ATC AGC TAC CAA GTG GAG AAG AAC AAG TAC GAC GCA 352
Val Leu Asp Trp Ile Ser Tyr Gln Val Glu Lys Asn Lys Tyr Asp Ala
 40 45 50
ACA GGG
Thr Gly
 358

```

## (2) INFORMATION FOR SEQ ID NO: 84:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 453 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Muscle

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 85..258
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.5  
seq LVSFVSSSEGTEQ/GE

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

```

AAGACCCTTT CCTGAGGTCC AGCAAGATAA TCCAGATCTC CAGTGGCAGA GAGTTGAGMN 60
TGATCCAGGA AAGTGAAGCA GGAG ATG CGG GAC TGC CCC GGG GTK GAA GBG 111
 Met Arg Asp Cys Pro Gly Val Glu Xaa
 -55 -50
ATC CTC GAC TGC TCT GMC AGG CAG AAG ACA GAA GGG TGC AGG CTT CAG 159
Ile Leu Asp Cys Ser Xaa Arg Gln Lys Thr Glu Gly Cys Arg Leu Gln
 -45 -40 -35
GCA GGA AAG GAG TGT GTG GAT TCT CCA GTG GAA GGA GGD CAG TCA GAA 207
Ala Gly Lys Glu Cys Val Asp Ser Pro Val Glu Gly Gly Gln Ser Glu
 -30 -25 -20
GCA CCT CCT TCT CTG GTA TCC TTT GCC GTC TCA TCA GAA GGC ACA GAG 255
Ala Pro Pro Ser Leu Val Ser Phe Ala Val Ser Ser Glu Gly Thr Glu
 -15 -10 -5
CAG GGA GAA GAT CCA CGC TCG GAA AAA GAT CAC AGC AGA CCT CAC AAG 303
Gln Gly Glu Asp Pro Arg Ser Glu Lys Asp His Ser Arg Pro His Lys
 1 5 10 15

```

|                                                                 |     |
|-----------------------------------------------------------------|-----|
| CAC CGA GCG CGG CAT GCA CGG CTC AGG AGG AGT GAA AGC CTG TCA GAM | 351 |
| His Arg Ala Arg His Ala Arg Leu Arg Arg Ser Glu Ser Leu Ser Xaa |     |
| 20 25 30                                                        |     |
| AAA CAA GTG AAG GAA GCA AAA TCT AMA TGC AAA AGC ATT GCC CTT CTT | 399 |
| Lys Gln Val Lys Glu Ala Lys Ser Xaa Cys Lys Ser Ile Ala Leu Leu |     |
| 35 40 45                                                        |     |
| CTA ACG GAT GCT CCC AAN CCC AAC TCC AAG GGG GTG TTG ATG TTT AAG | 447 |
| Leu Thr Asp Ala Pro Xaa Pro Asn Ser Lys Gly Val Leu Met Phe Lys |     |
| 50 55 60                                                        |     |
| AAG CGA                                                         | 453 |
| Lys Arg                                                         |     |
| 65                                                              |     |

## (2) INFORMATION FOR SEQ ID NO: 85:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 311 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Cancerous prostate

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 138..248
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.5  
seq LVFNFLILITILT/IW

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| AAGARTGCTT GTGAAGTAGC AACTAAAGTG GCAGTGTTC TTCTGAAATT CTCAGGCAGT  | 60  |
| CAGACTGTCT TAGGCAAATC TTGATAAAAT AGCCCTTATC CAGGTTTTTA TCTAAGGAAT | 120 |
| CCCAAGAAGA CTGGGGA ATG GAG AGA CAG TCA AGG GTT ATG TCA GAA AAG    | 170 |
| Met Glu Arg Gln Ser Arg Val Met Ser Glu Lys                       |     |
| -35 -30                                                           |     |
| GAT GAG TAT CAG TTT CAA CAT CAG GGA GCG GTG GAG CTG CTT GTC TTC   | 218 |
| Asp Glu Tyr Gln Phe Gln His Gln Gly Ala Val Glu Leu Leu Val Phe   |     |
| -25 -20 -15                                                       |     |
| AAT TTT TTG CTC ATC CTT ACC ATT TTG ACA ATC TGG TTA TTT AAA AAT   | 266 |
| Asn Phe Leu Leu Ile Leu Thr Ile Leu Thr Ile Trp Leu Phe Lys Asn   |     |
| -10 -5 1 5                                                        |     |
| CAT CGA TTC CGC TTC TTG CAT GAA ACT GGA GGA GCA ATG GTG TAT       | 311 |
| His Arg Phe Arg Phe Leu His Glu Thr Gly Gly Ala Met Val Tyr       |     |



10

15

20

## (2) INFORMATION FOR SEQ ID NO: 86:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 339 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Cancerous prostate

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 186..315
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 90..219  
id T70246  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 96..184
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..89  
id T70246  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 138..305
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 50..217  
id T70127  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 302..339
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 213..250  
id T70127  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 187..305
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 62..180  
id AA114263

est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 127..186  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 1..60  
id AA114263  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 302..339  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 176..213  
id AA114263  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 183..339  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 90  
region 73..229  
id T94480  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 183..339  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 90  
region 73..229  
id T89056  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 190..276  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 13.4  
seq SLLLVQLLTPCSA/QF

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

```
AATTTGCTTT CTCTTTTCC TTTCTCCGG ATGAGAGGCT AAGCCATART AGAAAGAATG 60
GAGAATTATT GATTGACCGT CTTTATWCTG TGGGCTCTGA TTCTCCAATG GGAATACCAA 120
GGGATGGTTT TCCATACTGG AACCCAAAGG TAAAGACACT CAAGGACAGA CATTTTTGGC 180
AGAGCATAG ATG AAA ATG GCA AGT TCC CTG GCT TTC CTT CTG CTC AAC TTT 231
 Met Lys Met Ala Ser Ser Leu Ala Phe Leu Leu Leu Asn Phe
 -25 -20

CAT GTC TCC CTC CTC TTG GTC CAG CTG CTC ACT CCT TGC TCA GCT CAG 279
His Val Ser Leu Leu Leu Val Gln Leu Leu Thr Pro Cys Ser Ala Gln
-15 -10 -5 1
```

TTT TCT GTG CTT KGA YCC TCT GGG CCC ATC CTG GCC ATG GTG GGT GAA 327  
 Phe Ser Val Leu Xaa Xaa Ser Gly Pro Ile Leu Ala Met Val Gly Glu  
                   5                  10                  15

GAC GCT GAT CTG 339  
 Asp Ala Asp Leu  
                   20

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 222 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Normal prostate

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 44..221
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
                           region 1..178  
                           id T27536  
                           est

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 100..195
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 12.6  
                           seq LLALLTVSTPSWC/QS

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

ATTTTTTCGG TCCTGGGGGA GCTAGGCCGG CGGCAGTGGT GGTGGCGGCG GCGCAAGGGT 60  
 GAKGGCGGCC CCAGAACCCC AGGTAGGTAG AGCAAGAAG ATG GTG TTT CTG CCC 114  
                                                           Met Val Phe Leu Pro  
                                                           -30  
 CTC AAA TGG TCC CTT GCA ACC ATG TCA TTT CTA CTT TCC TCA CTG TTG 162  
 Leu Lys Trp Ser Leu Ala Thr Met Ser Phe Leu Leu Ser Ser Leu Leu  
       -25                                  -20                                  -15  
 GCT CTC TTA ACT GTG TCC ACT CCT TCA TGG TGT CAG AGC ACT GAA GCA 210  
 Ala Leu Leu Thr Val Ser Thr Pro Ser Trp Cys Gln Ser Thr Glu Ala  
       -10                                  -5                                  1                                  5  
 TCC CCA AAA CGG 222  
 Ser Pro Lys Arg

## (2) INFORMATION FOR SEQ ID NO: 88:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 318 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Hypertrophic prostate

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 64..282
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 1..219  
id R93883  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 281..320
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 219..258  
id R93883  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 103..282
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 31..210  
id R84338  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 281..320
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 210..249  
id R84338  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 72..108
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..37  
id R84338  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 115..192
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 102..179  
id H38350  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 222..265
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 211..254  
id H38350  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 186..225
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 174..213  
id H38350  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 69..109
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 90  
region 54..94  
id H38350  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 102..142
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 66..106  
id AA010960  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 222..254
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 93  
region 191..223  
id AA010960  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 220..297
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 11.8  
seq SLLLLLLXC VHWS/QP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

```

AAGATTTCGT TTCCTGCATC TCCAAACATG GCGACCTAGG AGAAAGGGAA GAACAATTTT 60
TTCTCCTCTT TTGGGAAGGT TTGCGTCTAG TAGTGCCTGT GCCCCTGGGC AGATTGGAGA 120
GAAGAGGGAC GACTGGAGAA TCGTCGAGAA CCAGCGGAGA AAAGAAAAAG CAACGTTTAA 180
TTCTAGAAGG CCTCCTGTCC CTGCCTGCTC TGGGTGCTC ATG GAA TCA GCT GCT 234
 Met Glu Ser Ala Ala
 -25

GCC CTG CAC TTC TCC CGG CCA GCC TCC CTC CTC CTC CTS CTC CTC ASC 282
Ala Leu His Phe Ser Arg Pro Ala Ser Leu Leu Leu Leu Leu Leu Xaa
 -20 -15 -10

TGT GTG CAC TGG TCT CAG CCC AGT TTA TTG TCG TGG 318
Cys Val His Trp Ser Gln Pro Ser Leu Leu Ser Trp
 -5 1 5

```

## (2) INFORMATION FOR SEQ ID NO: 89:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 398 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Normal prostate

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 51..110
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 11.2  
seq AFLLLVALSYTLA/RD

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

```

AGAAGCTTGG ACCGCATCCT AGCCGCCGAC TCACACAAGG CAGAGTTGCC ATG GAR 56
 Met Glu
 -20

AAA ATT CCA GTG TCA GCA TTC TTG CTC CTT GTG GCC CTC TCC TAC ACT 104
Lys Ile Pro Val Ser Ala Phe Leu Leu Leu Val Ala Leu Ser Tyr Thr
 -15 -10 -5

CTG GCC AGA GAT ACC ACA GTC AAA CCT GGA GCC AAA AAG GAC ACA AAG 152
Leu Ala Arg Asp Thr Thr Val Lys Pro Gly Ala Lys Lys Asp Thr Lys
 1 5 10

GAC TCT CGA CCC AAA CTG CCC CAG ACC CTC TCC AGA GGT TGG GGT GAC 200
Asp Ser Arg Pro Lys Leu Pro Gln Thr Leu Ser Arg Gly Trp Gly Asp

```

| 15                                                              | 20 | 25 | 30 |     |
|-----------------------------------------------------------------|----|----|----|-----|
| CAA CTC ATC TGG ACT CAG ACA TAT GAA GAA GCT CTA TAT AAA TCC AAG |    |    |    | 248 |
| Gln Leu Ile Trp Thr Gln Thr Tyr Glu Glu Ala Leu Tyr Lys Ser Lys |    |    |    |     |
|                                                                 | 35 | 40 | 45 |     |
| ACA AGC AAC AAA CCC TTG ATG ATT ATT CAT CAC TTG GAT GAG TGC CCA |    |    |    | 296 |
| Thr Ser Asn Lys Pro Leu Met Ile Ile His His Leu Asp Glu Cys Pro |    |    |    |     |
|                                                                 | 50 | 55 | 60 |     |
| CAC AGT CAA GCT TTA AAG AAA GTG TTT GCT GAA AAT AAA GAA ATC CAG |    |    |    | 344 |
| His Ser Gln Ala Leu Lys Lys Val Phe Ala Glu Asn Lys Glu Ile Gln |    |    |    |     |
|                                                                 | 65 | 70 | 75 |     |
| AAA TTG GCA GAG CAG TTT GTC CTC CTC AAT CTG GTT TAT GAA ACA ACT |    |    |    | 392 |
| Lys Leu Ala Glu Gln Phe Val Leu Leu Asn Leu Val Tyr Glu Thr Thr |    |    |    |     |
|                                                                 | 80 | 85 | 90 |     |
| GAC AAA                                                         |    |    |    | 398 |
| Asp Lys                                                         |    |    |    |     |
| 95                                                              |    |    |    |     |

## (2) INFORMATION FOR SEQ ID NO: 90:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 292 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 47..289
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 3..245  
id H66924  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 77..214
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 10.3  
seq LVLLLVLTLLCSL/VP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| AASGCCGGAA GCGCGCGGAG ACCATGTAGT GAGACCCTCG CGAGGTCTGA GAGTCACTGG | 60  |
| AGCTACCAGA AGCATC ATG GGG CCC TGG GGA GAG CCA GAG CTC CTG GTG TGG | 112 |
| Met Gly Pro Trp Gly Glu Pro Glu Leu Leu Val Trp                   |     |

|                                                                 | -45 | -40 | -35 |     |
|-----------------------------------------------------------------|-----|-----|-----|-----|
| CGC CCC GAG GCG GTA GCT TCA GAG CCT CCA GTG CCT GTG GGG CTG GAG |     |     |     | 160 |
| Arg Pro Glu Ala Val Ala Ser Glu Pro Pro Val Pro Val Gly Leu Glu |     |     |     |     |
|                                                                 | -30 | -25 | -20 |     |
| GTG AAG TTG GGG GCC CTG GTG CTG CTG CTG GTG CTC ACC CTC CTC TGC |     |     |     | 208 |
| Val Lys Leu Gly Ala Leu Val Leu Leu Leu Val Leu Thr Leu Leu Cys |     |     |     |     |
|                                                                 | -15 | -10 | -5  |     |
| AGC CTG GTG CCC ATC TGT GTG CTG CGC CGG CCA GGA GCT AAC CAT GAA |     |     |     | 256 |
| Ser Leu Val Pro Ile Cys Val Leu Arg Arg Pro Gly Ala Asn His Glu |     |     |     |     |
|                                                                 | 1   | 5   | 10  |     |
| GGC TCA GCT TCC CGC CAG AAA GCC CTG AGC CCA AAG                 |     |     |     | 292 |
| Gly Ser Ala Ser Arg Gln Lys Ala Leu Ser Pro Lys                 |     |     |     |     |
|                                                                 | 15  | 20  | 25  |     |

## (2) INFORMATION FOR SEQ ID NO: 91:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 360 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Cancerous prostate

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 153..360
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 75..282  
id N29905  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 78..176
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 90  
region 1..99  
id N29905  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 153..360
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 75..282  
id N50844  
est



## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 78..176  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 90  
                           region 1..99  
                           id N50844  
                           est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 153..360  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 98  
                           region 75..282  
                           id N62597  
                           est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 153..360  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 98  
                           region 76..283  
                           id H03409  
                           est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 153..259  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 97  
                           region 76..182  
                           id R80247  
                           est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 7..54  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 10.1  
                           seq LLLQLAVLGAALA/AA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

|                                                                 |     |
|-----------------------------------------------------------------|-----|
| AGGAGA ATG GCT CCG CTT CTG TTG CAG CTG GCG GTG CTC GGC GCG GCG  | 48  |
| Met Ala Pro Leu Leu Leu Gln Leu Ala Val Leu Gly Ala Ala         |     |
| -15 -10 -5                                                      |     |
| CTG GCG GCC GCA GCC CTC GTA CTG ATT TCC ATC GTT GCA TTT ACA ACT | 96  |
| Leu Ala Ala Ala Ala Leu Val Leu Ile Ser Ile Val Ala Phe Thr Thr |     |
| 1 5 10                                                          |     |
| GCT ACA AAA ATG CCA GCA CTC CAT CGA CAT GAA GAA GAG AAA TTC TTC | 144 |
| Ala Thr Lys Met Pro Ala Leu His Arg His Glu Glu Glu Lys Phe Phe |     |
| 15 20 25 30                                                     |     |
| TTA AAT GCC AAA GGC CAG AAA GAA ACT TTA CCC AGC ATA TGG GAC TCA | 192 |
| Leu Asn Ala Lys Gly Gln Lys Glu Thr Leu Pro Ser Ile Trp Asp Ser |     |
| 35 40 45                                                        |     |

|                                                                 |     |
|-----------------------------------------------------------------|-----|
| CCT ACC AAA CAA CTT TCT GTC GTT GTG CCT TCA TAC AAT GAA GAA AAA | 240 |
| Pro Thr Lys Gln Leu Ser Val Val Val pro Ser Tyr Asn Glu Glu Lys |     |
| 50 55 60                                                        |     |
| CGG TTG CCT GTG ATG ATG GAT GAA GCT CTG AGC TAT CTA GAG AAG AGA | 288 |
| Arg Leu Pro Val Met Met Asp Glu Ala Leu Ser Tyr Leu Glu Lys Arg |     |
| 65 70 75                                                        |     |
| CAG AAA CGA GAT CCT GCG TTC ACT TAT GAA GTG ATA GTA GTT GAT GAT | 336 |
| Gln Lys Arg Asp Pro Ala Phe Thr Tyr Glu Val Ile Val Val Asp Asp |     |
| 80 85 90                                                        |     |
| GGC AGT AAA GAT CAG ACC TCA AAG                                 | 360 |
| Gly Ser Lys Asp Gln Thr Ser Lys                                 |     |
| 95 100                                                          |     |

## (2) INFORMATION FOR SEQ ID NO: 92:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 451 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymphocytes

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 338..453
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..116  
id R09346  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 338..453
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..116  
id R06965  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 71..151
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 9.8  
seq SALLVGFLSVIFA/LV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

```

AACTACCCAG AGSACTGCCG CCGCCTCTCC AAGTTCTTGT GGCCCCCGCG GTGCGSAGTA 60
TGGGGCGCTG ATG GCC ATG GAG GGC TAC TGG CGC TTC CTR RCG CTG CTG 109
 Met Ala Met Glu Gly Tyr Trp Arg Phe Leu Xaa Leu Leu
 -25 -20 -15

GGG TCG GCA CTG CTC GTC GGC TTC CTG TCG GTG ATC TTC GCC CTC GTC 157
Gly Ser Ala Leu Leu Val Gly Phe Leu Ser Val Ile Phe Ala Leu Val
 -10 -5 1

TGG GTC CTC CAC TAC CGA GAG GGG CTT GGC TGG GAT GGG AGC GCA CTA 205
Trp Val Leu His Tyr Arg Glu Gly Leu Gly Trp Asp Gly Ser Ala Leu
 5 10 15

GAG TTT AAC TGG SRC CCA GTG CTC ATG GTC ACC GGC TTC GTC TTC ATC 253
Glu Phe Asn Trp Xaa Pro Val Leu Met Val Thr Gly Phe Val Phe Ile
 20 25 30

CAG GGC ATC GCC ATC ATC GTC TAC AGA CTG CCG TGG ACC TGG AAA TGC 301
Gln Gly Ile Ala Ile Ile Val Tyr Arg Leu Pro Trp Thr Trp Lys Cys
 35 40 45 50

AGC AAG CTC CTG ATG AAA TCC ATC CAT GCA RGG TTA AAT GCA GTT GCT 349
Ser Lys Leu Leu Met Lys Ser Ile His Ala Xaa Leu Asn Ala Val Ala
 55 60 65

GCC ATT CTT GCA ATT ATC TCT GTG GTG GCC GTG TTT GAG AAC CAC AAT 397
Ala Ile Leu Ala Ile Ile Ser Val Val Ala Val Phe Glu Asn His Asn
 70 75 80

GTT AAC AAT ATA GCC AAT ATG TAC AGT CTG CAC AGC TGG GTT GGA CTG 445
Val Asn Asn Ile Ala Asn Met Tyr Ser Leu His Ser Trp Val Gly Leu
 85 90 95

ATA GCT 451
Ile Ala
100

```

## (2) INFORMATION FOR SEQ ID NO: 93:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 458 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 114..376
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 36..298  
id W17274

est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 371..459
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 292..380  
id W17274  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 78..120
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..43  
id W17274  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 96..289
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 29..222  
id AA149456  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 382..459
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 317..394  
id AA149456  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 292..367
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 224..299  
id AA149456  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 153..398
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 2..247  
id W67885  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 381..424
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 93

region 231..274  
id W67885  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 414..443  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 90  
region 265..294  
id W67885  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 72..122  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 9.3  
seq LALSLLILVLAFG/IP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

```

AACAGACCCC CAACTGCGAG CTGCCCACCN CACCCTCAGC TCTGGCCTCT TACTCACCTT 60
CTACCACAGA C ATG GCT CAG TCA CTG GCT CTG AGC CTC CTT ATC CTG GTT 110
 Met Ala Gln Ser Leu Ala Leu Ser Leu Leu Ile Leu Val
 -15 -10 -5
CTG GCC TTT GGC ATC CCC AGG ACC CAA GGC AGT GAT GGA GGG GCT CAG 158
Leu Ala Phe Gly Ile Pro Arg Thr Gln Gly Ser Asp Gly Gly Ala Gln
 1 5 10
GAC TGT TGC CTC AAG TAC AGC CAA AGG AAG ATT CCC GCC AAG GTT GTC 206
Asp Cys Cys Leu Lys Tyr Ser Gln Arg Lys Ile Pro Ala Lys Val Val
 15 20 25
CGC AGC TAC CGG AAG CAG GAA CCA AGC TTA GGC TGC TCC ATC CCA GCT 254
Arg Ser Tyr Arg Lys Gln Glu Pro Ser Leu Gly Cys Ser Ile Pro Ala
 30 35 40
ATC CTG TTC TTG CCC CGC AAG CGC TCT CAG GCA GAG CTA TGT GCA GAC 302
Ile Leu Phe Leu Pro Arg Lys Arg Ser Gln Ala Glu Leu Cys Ala Asp
 45 50 55 60
CCA AAG GAG CTC TGG GTG CAG CAG CTG ATG CAG CAT CTG GAC AAG ACA 350
Pro Lys Glu Leu Trp Val Gln Gln Leu Met Gln His Leu Asp Lys Thr
 65 70 75
CCA TCC CCA CAG AAA CCA GCC CAG GGC TGC AGG AAG GAC AGG GGG GCC 398
Pro Ser Pro Gln Lys Pro Ala Gln Gly Cys Arg Lys Asp Arg Gly Ala
 80 85 90
TCC AAG ACT GGC AAG AAA GGA AAR GGC TCC AAA GGC TGC AAG AGG ACT 446
Ser Lys Thr Gly Lys Lys Gly Lys Gly Ser Lys Gly Cys Lys Arg Thr
 95 100 105
GAG CGG TCA CAG 458
Glu Arg Ser Gln
 110

```

## (2) INFORMATION FOR SEQ ID NO: 94:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 186 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Brain

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 52..184
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 1..133  
id W93799  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 19..63
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 8.4  
seq AMWLLCVALAVLA/WG

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

```

AAGTGCTGCT TACCCATC ATG GAA GCA ATG TGG CTC CTG TGT GTG GCG TTG 51
 Met Glu Ala Met Trp Leu Leu Cys Val Ala Leu
 -15 -10 -5

GCG GTC TTG GCA TGG GGC TTC CTC TGG GTT TGG GAC TCC TCA GAA CGA 99
Ala Val Leu Ala Trp Gly Phe Leu Trp Val Trp Asp Ser Ser Glu Arg
 1 5 10

ATG AAG AGT CGG GAG CAG GGA RGA CGG CTG GGA GCC GAA AGC CGG ACC 147
Met Lys Ser Arg Glu Gln Gly Xaa Arg Leu Gly Ala Glu Ser Arg Thr
 15 20 25

CTG CTG GTC ATA GCG CAC CCT GAC GAT GAA GCC ATG TGG 186
Leu Leu Val Ile Ala His Pro Asp Asp Glu Ala Met Trp
 30 35 40

```

## (2) INFORMATION FOR SEQ ID NO: 95:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 427 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Brain

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 266..427

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 99  
region 137..298  
id AA081755  
est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 129..267

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100  
region 1..139  
id AA081755  
est

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide

(B) LOCATION: 212..325

(C) IDENTIFICATION METHOD: Von Heijne matrix

(D) OTHER INFORMATION: score 7.8  
seq LVFTVSLFAWICC/QR

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| AAAGAAGAGC CAAAACAGGA ACCGAGGTGG CAAATCACTG TGCAGGGGCG AGTGGACCTC | 60  |
| CCTCTTTGCC TCCTCCCTGT TCCAGGAGCT GGTGCCCTGG GCTCTGCGCT GTTGTTTTCA | 120 |
| GCGCTCCGAA AGCCGGCGCT TGAGATCCAG GCAAGTGAAT CCAGCCAGGC AGTTTTCCTT | 180 |
| TCAGCACCTC GGACAGAACA CGCAGTAAAA A ATG GCT CCG ATC ACC ACC AGC    | 232 |
| Met Ala Pro Ile Thr Thr Ser                                       |     |
| -35                                                               |     |
| CGG GAA GAA TTT GAT GAA ATC CCC ACA GTG GTG GGG ATC TTC AGT GCA   | 280 |
| Arg Glu Glu Phe Asp Glu Ile Pro Thr Val Val Gly Ile Phe Ser Ala   |     |
| -30 -25 -20                                                       |     |
| TTT GGC CTG GTC TTC ACA GTC TCT CTC TTT GCA TGG ATC TGC TGT CAG   | 328 |
| Phe Gly Leu Val Phe Thr Val Ser Leu Phe Ala Trp Ile Cys Cys Gln   |     |
| -15 -10 -5 1                                                      |     |
| AGA AAA TCA TCC AAG TCT AAC AAG ACT CCT CCA TAC AAG TTT GTG CAT   | 376 |
| Arg Lys Ser Ser Lys Ser Asn Lys Thr Pro Pro Tyr Lys Phe Val His   |     |
| 5 10 15                                                           |     |
| GTG CTT WAG GGA GTT GAT ATT TAC CCT GAA AAC CTA AAT AGC AAA AAG   | 424 |
| Val Leu Xaa Gly Val Asp Ile Tyr Pro Glu Asn Leu Asn Ser Lys Lys   |     |
| 20 25 30                                                          |     |
| AAG                                                               | 427 |

(2) INFORMATION FOR SEQ ID NO: 96:

(A) LENGTH: 400 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Umbilical cord

[illegible]

(A) NAME/KEY: other  
(B) LOCATION: 226..307  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 95  
region 8..89  
id T53693  
est

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 38..91  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 7.4  
seq GWLVLCVLAISLA/SM

|                                                                  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |     |
|------------------------------------------------------------------|--|--|--|--|--|--|--|--|--|--|--|--|--|--|-----|
| AATCCAGTYG GASTTGACAA CAGGAGGCAG AGGCATC ATG GAG GGT CCC CGG GGA |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 55  |
| Met Glu Gly Pro Arg Gly                                          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |     |
| -15                                                              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |     |
| TGG CTG GTG CTC TGT GTG CTG GCC ATA TCG CTG GCC TCT ATG GTG ACC  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 103 |
| Trp Leu Val Leu Cys Val Leu Ala Ile Ser Leu Ala Ser Met Val Thr  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |     |
| -10 -5 1                                                         |  |  |  |  |  |  |  |  |  |  |  |  |  |  |     |
| GAG GAC TTG TGC CGA GCA CCA GAC GGG AAG AAA GGG GAG GCA GGA AGA  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 151 |
| Glu Asp Leu Cys Arg Ala Pro Asp Gly Lys Lys Gly Glu Ala Gly Arg  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |     |
| 5 10 15 20                                                       |  |  |  |  |  |  |  |  |  |  |  |  |  |  |     |
| CCT GGC AGA CGG GGG CGG CCA GGC CTC AAG GGG GAG CAA GGG GAG CCG  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 199 |
| Pro Gly Arg Arg Gly Arg Pro Gly Leu Lys Gly Glu Gln Gly Glu Pro  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |     |
| 25 30 35                                                         |  |  |  |  |  |  |  |  |  |  |  |  |  |  |     |



```

GGG GCC CCT GGC ATC CGG ACA GGC ATC CAA GGC CTT AAA GGA GAC CAG 247
Gly Ala Pro Gly Ile Arg Thr Gly Ile Gln Gly Leu Lys Gly Asp Gln
 40 45 50

GGG GAA CCT GGG CCC TCT GGA AAC CCC GGC AAG GTG GGC TAC CCA GGG 295
Gly Glu Pro Gly Pro Ser Gly Asn Pro Gly Lys Val Gly Tyr Pro Gly
 55 60 65

CCC AGC GGC CCC CTC GGA GCC CGT GGC ATC CCG GGA ATT AAA GGC ACC 343
Pro Ser Gly Pro Leu Gly Ala Arg Gly Ile Pro Gly Ile Lys Gly Thr
 70 75 80

AAG GGC AGC CCA GGA AAC ATC AAG GAC CAG CCG AGG CCA GCC TTC TCC 391
Lys Gly Ser Pro Gly Asn Ile Lys Asp Gln Pro Arg Pro Ala Phe Ser
 85 90 95 100

GCC ATT CGG 400
Ala Ile Arg

```

## (2) INFORMATION FOR SEQ ID NO: 97:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 288 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Muscle

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 42..132
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 1..91  
id N77056  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 52..240
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 7.2  
seq VLLTLLLIAFIFL/II

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

```

AAGTCTTAGA CGACTGCGTC GTGCTATGAC CGGACTTTTT CTTGAAAGGG G ATG ACA 57
 Met Thr

GCA TGG GAG GCA ATG GCT CCA CAT GTA AAC CCG ACA CTG AAA GAC AAG 105
Ala Trp Glu Ala Met Ala Pro His Val Asn Pro Thr Leu Lys Asp Lys
-60 -55 -50

```

```

GCA CTC TCT CCA CAG CAG SCC CMA CMA ACT AGC CCT GCA CCC TGT CNY 153
Ala Leu Ser Pro Gln Gln Xaa Xaa Xaa Thr Ser Pro Ala Pro Cys Xaa
-45 -40 -35 -30

TCT AAC CAC CAC AAC AAA AAA CAT TTA ATC CTT GCC TTT TGT GCT GGG 201
Ser Asn His His Asn Lys Lys His Leu Ile Leu Ala Phe Cys Ala Gly
-25 -20 -15

GTT CTA CTG ACA CTG CTG CTG ATA GCC TTT ATC TTC CTC ATC ATA AAG 249
Val Leu Leu Thr Leu Leu Leu Ile Ala Phe Ile Phe Leu Ile Ile Lys
-10 -5 1

AGC TAC AGA AAA TAT CAC TCC AAG CCC CAG GCC CCC GGG 288
Ser Tyr Arg Lys Tyr His Ser Lys Pro Gln Ala Pro Gly
5 10 15

```

## (2) INFORMATION FOR SEQ ID NO: 98:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 333 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (D) DEVELOPMENTAL STAGE: Fetal
- (F) TISSUE TYPE: kidney

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 211..313
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 2..104  
id N57441  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 136..189
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 7.1  
seq LLCECLLLXAGYA/HD

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

```

GAACAATTCG ATGACGAGGC CCAGGAAGCA CGCTGAAACC CTGGGCGGCG GCAAGCTGTG 60
CGACCTCTTC TCGGCCCGGC CTGGGCAGGT GTCTTCCTCG AGAGGCAGGC AGGGGATCBC 120
GGACCCTTAT ACAGG ATG CTG TGT TCT TTG CTC CTT TGT GAA TGT CTG TTG 171
Met Leu Cys Ser Leu Leu Leu Cys Glu Cys Leu Leu
-15 -10

```

|                                                                 |     |
|-----------------------------------------------------------------|-----|
| CTG GYN GCT GGT TAT GCT CAT GAT GAT GAC TGG ATT GAC CCC ACA GAC | 219 |
| Leu Xaa Ala Gly Tyr Ala His Asp Asp Asp Trp Ile Asp Pro Thr Asp |     |
| -5 1 5 10                                                       |     |
| ATG CTT AAC TAT GAT GCT GCT TCA GGA ACA ATG AGA AAA TCT CAG GCA | 267 |
| Met Leu Asn Tyr Asp Ala Ala Ser Gly Thr Met Arg Lys Ser Gln Ala |     |
| 15 20 25                                                        |     |
| AAA TAT GGT ATT TCA GGG GAA AAG GAT GTC AGT CCT GAC TTG TCA TGT | 315 |
| Lys Tyr Gly Ile Ser Gly Glu Lys Asp Val Ser Pro Asp Leu Ser Cys |     |
| 30 35 40                                                        |     |
| GCT GRT GAA ATA TCA GAA                                         | 333 |
| Ala Xaa Glu Ile Ser Glu                                         |     |
| 45                                                              |     |

## (2) INFORMATION FOR SEQ ID NO: 99:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 462 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (D) DEVELOPMENTAL STAGE: Fetal
- (F) TISSUE TYPE: kidney

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 158..307
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 129..278  
id R18809  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 99..157
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 71..129  
id R18809  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 323..371
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 299..347  
id R18809  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 305..441  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 96  
region 141..277  
id R88070  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 167..300  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 1..134  
id R88070  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 158..307  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 96  
region 68..217  
id T85919  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 98..157  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98  
region 9..68  
id T85919  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 158..317  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 93  
region 129..288  
id R60434  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 99..157  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98  
region 71..129  
id R60434  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 158..307  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 96  
region 86..235  
id W23910  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 98..157  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 96  
                           region 27..86  
                           id W23910  
                           est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 325..381  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 6.9  
                           seq LVXSLPVHCLTFA/SS

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

```

AAGTTGGTGG AGTTCTGCCC GGATGGAAGC TCCGCCCGCG GAGTGATGGT GGCCTCAGCG 60
AAGATGGGCC GGGCAGGGAC CATGGCGGTG GCAGCAGAGC TTCGAGAGCT GTGCCCAGGA 120
GTGAACAACC AGCCCTACCT CTGTGAGAGT KGTCACCTGC TCGGGGAAM CTGGCTGCTG 180
CACCTACTAC TATGAGCTCT GGTGGTTCTG GCTGCTCTGG ACTGTCCTCA TCCTCTTTAG 240
CTGCTGTTGC GCCTTCGCC ACCGACGAGC TAAACTCAGG CTGCAACAAC AGCAGCGGCA 300
SSTGAAACAA CTTGTTGGCC TATC ATG GGG CAT GCC ATG GGG CTG GTN STT 351
 Met Gly His Ala Met Gly Leu Val Xaa
 -15

TCC CTA CCG GTT CAC TGC TTG ACC TTC GCT TCC TCA GCA CCT TCA AGC 399
Ser Leu Pro Val His Cys Leu Thr Phe Ala Ser Ser Ala Pro Ser Ser
-10 -5 1 5

CCC CAG CCT ACG AGG ATG TGG TTC AMC GCC CAG GCA CAC CAM CCC CCC 447
Pro Gln Pro Thr Arg Met Trp Phe Xaa Ala Gln Ala His Xaa Pro Pro
 10 15 20

CTT ATA CTG GGC CCG 462
Leu Ile Leu Gly Pro
 25

```

## (2) INFORMATION FOR SEQ ID NO: 100:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 451 base pairs  
 (B) TYPE: NUCLEIC ACID  
 (C) STRANDEDNESS: DOUBLE  
 (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
 (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 156..288  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 1..133  
id AA081350  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 289..396  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 133..240  
id AA081350  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 422..453  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 90  
region 269..300  
id AA081350  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 289..453  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 67..231  
id AA046671  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 222..289  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 1..68  
id AA046671  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 104..151  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 6.7  
seq CFSLVLLLSIWT/TR

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

AATAGTTCCA GAACTCTCCA TCCGGACTAG TTATTGAGCA TCTGCCTCTC ATATCACCAG 60

TGGCCATCTG AGGTGTTTCC CTGGCTCTGA AGGGGTAGGC ACG ATG GCC AGG TGC 115

Met Ala Arg Cys

-15

|                                                                 |     |
|-----------------------------------------------------------------|-----|
| TTC AGC CTG GTG TTG CTT CTC ACT TCC ATC TGG ACC ACG AGG CTC CTG | 163 |
| Phe Ser Leu Val Leu Leu Leu Thr Ser Ile Trp Thr Thr Arg Leu Leu |     |
| -10 -5 1                                                        |     |
| GTC CAA GGC TCT TTG CGT GCA GAA GAG CTT TCC ATC CAG GTG TCA TGC | 211 |
| Val Gln Gly Ser Leu Arg Ala Glu Glu Leu Ser Ile Gln Val Ser Cys |     |
| 5 10 15 20                                                      |     |
| AGA ATT ATG GGG ATC ACC CTT GTG AGC AAA AAG GCG AAC CAG CAG CTG | 259 |
| Arg Ile Met Gly Ile Thr Leu Val Ser Lys Lys Ala Asn Gln Gln Leu |     |
| 25 30 35                                                        |     |
| AAT TTC ACA GAA GCT AAG GAG GCC TGT AGG CTG CTG GGA CTA AGT TTG | 307 |
| Asn Phe Thr Glu Ala Lys Glu Ala Cys Arg Leu Leu Gly Leu Ser Leu |     |
| 40 45 50                                                        |     |
| GCC GGC AAG GAC CAA GTT GAA ACA GCC TTG AAA GCT AGC TTT GAA ACT | 355 |
| Ala Gly Lys Asp Gln Val Glu Thr Ala Leu Lys Ala Ser Phe Glu Thr |     |
| 55 60 65                                                        |     |
| TGC AGC TAT GGC TGG GTT GGA GAT GGA TTC GTG GTC ATC TCT AGG ATT | 403 |
| Cys Ser Tyr Gly Trp Val Gly Asp Gly Phe Val Val Ile Ser Arg Ile |     |
| 70 75 80                                                        |     |
| AGC CCA AAC CCC AAG TGT GGG AAA AAT GGG GTG GGT GTC CTG ATT TGG | 451 |
| Ser Pro Asn Pro Lys Cys Gly Lys Asn Gly Val Gly Val Leu Ile Trp |     |
| 85 90 95 100                                                    |     |

## (2) INFORMATION FOR SEQ ID NO: 101:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 369 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (D) DEVELOPMENTAL STAGE: Fetal
- (F) TISSUE TYPE: kidney

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 67..366
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 2..301  
id AA056199  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 152..366
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..215

id R66275  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 117..221  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 99..203  
id AA054476  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 39..120  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 96  
region 22..103  
id AA054476  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 232..366  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 1..135  
id AA143025  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 242..366  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 84..208  
id W90481  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 175..351  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 6.6  
seq VLAQLAFLSQISQ/CI

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

```
ACTTTTCCGG CTGACTTCTG AGAAGGTTGC GCASAGCTGT GCCCGGCAGT CTAGAGGCGC 60
AGAAGAGGAA GCCATCGCCT GGCCCCGGCT CTCTGGACCT TGTCTCGCTC GGGAGCGGAA 120
ACAGCGGCAG CCAGAGAACT GTTTTAATCA TGGACAAACA AAATCACAG ATGA ATG 177
 Met
CTT CTC ACC CGG AAA CAA ACT TGC CAG TTG GGT ATC CTC CTC AGT ATC 225
Leu Leu Thr Arg Lys Gln Thr Cys Gln Leu Gly Ile Leu Leu Ser Ile
 -55 -50 -45
CAC CGA CAG CAT TCC AAG GAC CTC CAG GAT ATA GTG GCT ACC CTG GGC 273
```



His Arg Gln His Ser Lys Asp Leu Gln Asp Ile Val Ala Thr Leu Gly  
-40 -35 -30

CCC AGG TCA GCT ACC CAC CCC CAC CAG CCG GCC ATT CAG GTC CTG GCC 321  
Pro Arg Ser Ala Thr His Pro His Gln Pro Ala Ile Gln Val Leu Ala  
-25 -20 -15

CAG CTG GCT TTC CTG TCC CAA ATC AGC CAG TGT ATA ATC AGC CAG CGG 369  
Gln Leu Ala Phe Leu Ser Gln Ile Ser Gln Cys Ile Ile Ser Gln Arg  
-10 -5 1 5

## (2) INFORMATION FOR SEQ ID NO: 102:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 414 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Cancerous prostate

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 286..414
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 211..339  
id AA284366  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 166..300
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 92  
region 92..226  
id AA284366  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 72..177
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 1..106  
id AA284366  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 199..282
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 6.6  
seq IVSLLGFVATVTL/IP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

```

AGAACATAGG TTGCCTTAGA GAGGTTCCCC GGTGTCCCGA CGGCGGCTCA AGTCAGAGTT 60
GCTGGGTTTT GCTCAGATTG GTGTGGGAAG AGCCTGCCTG TGGGGAGCGG CCACTCCATA 120
CTGCTGARGC CTCAGGACTG CTGCTCAGCT TGCCCGTTAC CTGAAGAGGC GGCGGAGCGG 180
NGCCCCTGAC CGGTCACC ATG TGG GCC TTC TCG GAA TTG CCC ATG CCG CTG 231
 Met Trp Ala Phe Ser Glu Leu Pro Met Pro Leu
 -25 -20

CTG ATC AAT TTG ATC GTC TCG CTG CTG GGA TTT GTG GCC ACA GTC ACC 279
Leu Ile Asn Leu Ile Val Ser Leu Leu Gly Phe Val Ala Thr Val Thr
 -15 -10 -5

CTC ATC CCG GCC TTC CGG GGC CAC TTC ATT GCT GCG CGC CTC TGT GGT 327
Leu Ile Pro Ala Phe Arg Gly His Phe Ile Ala Ala Arg Leu Cys Gly
 1 5 10 15

CAG GAC CTC AAC AAA ACC AGC CGA CAG CAG ATC CCA GAA TCC CAG GGA 375
Gln Asp Leu Asn Lys Thr Ser Arg Gln Gln Ile Pro Glu Ser Gln Gly
 20 25 30

GTG ATC AGC GGT GCT GTT TTC CTT ATC ATC CTC TTC TGC 414
Val Ile Ser Gly Ala Val Phe Leu Ile Ile Leu Phe Cys
 35 40

```

(2) INFORMATION FOR SEQ ID NO: 103:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 457 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Hypertrophic prostate

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 209..341
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 241..373  
id H87867  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 28..124
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 63..159  
id H87867

est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 168..207
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 92  
region 201..240  
id H87867  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 224..459
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 1..236  
id N87591  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 263..453
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 65..255  
id AA172091  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 202..251
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 4..53  
id AA172091  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 263..459
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 38..234  
id H85080  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 225..261
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..37  
id H85080  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 212..280
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.4

seq PASLSLLTFKVYA/AP

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

```

GACGCGCTGC GGCTCAGCGA CGCGGCTTCT AGAACCGGGT GATTGAACTA AACCTTCGCC 60
GCACCGAGTT TGCAGTACGG CCGTCACCCG CACCGCTGCC TGCTTGCGGT TGGAGAAATC 120
AARGGGCCCT ACCGGGCCTC CGTAGTCACC TCTCTATAGT GGGCGTGGCC GAGGCCGGGG 180
TGACCCTGCC GGAGCCTCCG CTGCCAGCGA C ATG TTC AAG GTA ATT CAG AGG 232
 Met Phe Lys Val Ile Gln Arg
 -20

TCC GTG GGG CCA GCC AGC CTG AGC TTG CTC ACC TTC AAA GTC TAT GCA 280
Ser Val Gly Pro Ala Ser Leu Ser Leu Leu Thr Phe Lys Val Tyr Ala
-15 -10 -5

GCA CCA AAA AAG GAC TCA CCT CCC AAA AAT TCC GTG AAG GTT GAT GAG 328
Ala Pro Lys Lys Asp Ser Pro Pro Lys Asn Ser Val Lys Val Asp Glu
 1 5 10 15

CTT TCA CTC TAC TCA GTT CCT GAG GGT CAA TCG AAG TAT GTG GAG GAG 376
Leu Ser Leu Tyr Ser Val Pro Glu Gly Gln Ser Lys Tyr Val Glu Glu
 20 25 30

GCA AGG AGC CAG CTT GAA GAA AGC ATC TCA CAG CTC CGA CAC TAT TGC 424
Ala Arg Ser Gln Leu Glu Glu Ser Ile Ser Gln Leu Arg His Tyr Cys
 35 40 45

GAG CCA TAC ACA ACC TGG TGT CAG GAA ACG TAC 457
Glu Pro Tyr Thr Thr Trp Cys Gln Glu Thr Tyr
 50 55

```

(2) INFORMATION FOR SEQ ID NO: 104:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 439 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (D) DEVELOPMENTAL STAGE: Fetal
- (F) TISSUE TYPE: kidney

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 141..354
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 38..251  
id T94226  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 225..373
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..149  
id W95280  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 371..437
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 148..214  
id W95280  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 167..289
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 91  
region 2..124  
id N55978  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 262..326
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 98..162  
id N55978  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 379..437
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 270..328  
id N55978  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 317..373
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 154..210  
id N55978  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 20..427
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 5.4  
seq LISVALVQGWALG/GG

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

```

AACCGTGGCC TGCGACGAA ATG GCG AAA AGT CTT TTG AAG ACA GCC TCT CTG 52
 Met Ala Lys Ser Leu Leu Lys Thr Ala Ser Leu
 -135 -130

TCT GGA AGG ACA AAA TTG CTA CAT CAA ACA GGA TTG TCA CTT TAT AGT 100
Ser Gly Arg Thr Lys Leu Leu His Gln Thr Gly Leu Ser Leu Tyr Ser
-125 -120 -115 -110

ACA TCC CAT GGA TTT TAT GAG GAA GAA GTG AAA AAA ACA CTT CAG CAG 148
Thr Ser His Gly Phe Tyr Glu Glu Glu Val Lys Lys Thr Leu Gln Gln
 -105 -100 -95

TTT CCT GGT GGA TCC ATT GAC CTT CAG AAG GAA GAC AAT GGC ATT GGC 196
Phe Pro Gly Gly Ser Ile Asp Leu Gln Lys Glu Asp Asn Gly Ile Gly
 -90 -85 -80

ATT CTT ACT CTG AAC AAT CCA AGT AGA ATG AAT GCC TTT TCA GGT GTT 244
Ile Leu Thr Leu Asn Asn Pro Ser Arg Met Asn Ala Phe Ser Gly Val
 -75 -70 -65

ATG ATG CTA CAA CTT CTG GAA AAA GTA ATT GAA TTG GAA AAT TGG ACA 292
Met Met Leu Gln Leu Leu Glu Lys Val Ile Glu Leu Glu Asn Trp Thr
 -60 -55 -50

GAG GGG AAA GGC CTC ATT GTC CGT GGG GCA AAA AAT ACT TTC TCT TCA 340
Glu Gly Lys Gly Leu Ile Val Arg Gly Ala Lys Asn Thr Phe Ser Ser
-45 -40 -35 -30

GGA TCT GAT CTG AAT GCT GTG AAA TCA CTA GGA CTC CAG AGA CTT CCT 388
Gly Ser Asp Leu Asn Ala Val Lys Ser Leu Gly Leu Gln Arg Leu Pro
 -25 -20 -15

TTA ATA AGT GTT GCG CTG GTT CAA GGT TGG GCA TTG GGT GGA GGA GCA 436
Leu Ile Ser Val Ala Leu Val Gln Gly Trp Ala Leu Gly Gly Gly Ala
 -10 -5 1

GCG
Ala
439

```

(2) INFORMATION FOR SEQ ID NO: 105:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 323 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Substantia nigra

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 116..212

(C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 95  
                           region 125..221  
                           id HUMEST2D1  
                           est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 214..322  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 99  
                           region 9..117  
                           id AA115085  
                           est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 132..263  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 5.3  
                           seq PLLKILHAAGAQG/EM

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

```

AATTTCAVVA TGCTGCCGAG GCCCTAGGAT CTGTGACTGC CACCCCTCCC CCCACCCGGG 60
CTCGGCGGGG GAGCGACTCA TGGAGCTGCC GTAAGTTTTA CCAACAGACT GCAGTTTCTT 120
TCACTACCAA A ATG ACA TCA TTT TCC ACC TCT GCT CAG TGT TCA ACA TCT 170
 Met Thr Ser Phe Ser Thr Ser Ala Gln Cys Ser Thr Ser
 -40 -35

GAC AGT GCT TGC AGG ATC TCT CCT GGA CAA ATC AAT SVG GTA CGA CCA 218
Asp Ser Ala Cys Arg Ile Ser Pro Gly Gln Ile Asn Xaa Val Arg Pro
-30 -25 -20

AAA CTG CCG CTT TTG AAG ATT TTG CAT GCA GCA GGT GCG CAA GGT GAA 266
Lys Leu Pro Leu Leu Lys Ile Leu His Ala Ala Gly Ala Gln Gly Glu
-15 -10 -5 1

ATG TTC ACT GTT AAA GAG GTC ATG CAC TAT TTA GGT CAG TAC ATA ATG 314
Met Phe Thr Val Lys Glu Val Met His Tyr Leu Gly Gln Tyr Ile Met
 5 10 15

GTG AAG CAG 323
Val Lys Gln
 20

```

## (2) INFORMATION FOR SEQ ID NO: 106:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 478 base pairs  
 (B) TYPE: NUCLEIC ACID  
 (C) STRANDEDNESS: DOUBLE  
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: cDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(F) TISSUE TYPE: Lung (cells)

## (ix) FEATURE:

- (A) NAME/KEY: other  
(B) LOCATION: 104..370  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 1..267  
id AA114062  
est

## (ix) FEATURE:

- (A) NAME/KEY: other  
(B) LOCATION: 409..451  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 93  
region 309..351  
id AA114062  
est

## (ix) FEATURE:

- (A) NAME/KEY: other  
(B) LOCATION: 388..420  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 90  
region 287..319  
id AA114062  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 5..340  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 5.1  
seq AFAWLGVVPLTAC/RI

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

|                                                                  |     |
|------------------------------------------------------------------|-----|
| AAAG ATG GAC ACC GCG GAG GAA GAC ATA TGT AGA GTG TGT CGG TCA GAA | 49  |
| Met Asp Thr Ala Glu Glu Asp Ile Cys Arg Val Cys Arg Ser Glu      |     |
| -110 -105 -100                                                   |     |
| GGA ACA CCT GAG AAA CCG CTT TAT CAT CCT TGT GTA TGT ACT GGC AGT  | 97  |
| Gly Thr Pro Glu Lys Pro Leu Tyr His Pro Cys Val Cys Thr Gly Ser  |     |
| -95 -90 -85                                                      |     |
| ATT AAG TTN GTC CAT CAA GAA TGC TTA GTT CAA TGG CTG AAA CAC AGT  | 145 |
| Ile Lys Xaa Val His Gln Glu Cys Leu Val Gln Trp Leu Lys His Ser  |     |
| -80 -75 -70                                                      |     |
| CGA AAA GAA TAC TGT GAA TTA TGC AAG CAC AGA TTT GCT TTT ACA CCA  | 193 |
| Arg Lys Glu Tyr Cys Glu Leu Cys Lys His Arg Phe Ala Phe Thr Pro  |     |
| -65 -60 -55 -50                                                  |     |
| ATT TAT TCT CCA GAT ATG CCT TCA CGG CTT CCA ATT CAA GAC ATA TTT  | 241 |
| Ile Tyr Ser Pro Asp Met Pro Ser Arg Leu Pro Ile Gln Asp Ile Phe  |     |
| -45 -40 -35                                                      |     |



|                                                                 |     |
|-----------------------------------------------------------------|-----|
| GCT GGA CTG GTT ACA AGT ATT GGC ACT GCA ATA CGA TAT TGG TTT CAT | 289 |
| Ala Gly Leu Val Thr Ser Ile Gly Thr Ala Ile Arg Tyr Trp Phe His |     |
| -30 -25 -20                                                     |     |
| TAT ACA CTT GTG GCC TTT GCA TGG TTG GGA GTT GTT CCT CTT ACA GCA | 337 |
| Tyr Thr Leu Val Ala Phe Ala Trp Leu Gly Val Val Pro Leu Thr Ala |     |
| -15 -10 -5                                                      |     |
| TGC CGC ATC TAC AAG TGC TTG TTT ACT GGC TCC GTG AGC TCA CTA CTG | 385 |
| Cys Arg Ile Tyr Lys Cys Leu Phe Thr Gly Ser Val Ser Ser Leu Leu |     |
| 1 5 10 15                                                       |     |
| ACG CTG CCA TTA GAT ATG CTG TCA ACG GAA AAT TTG TTG GCA GAT TGT | 433 |
| Thr Leu Pro Leu Asp Met Leu Ser Thr Glu Asn Leu Leu Ala Asp Cys |     |
| 20 25 30                                                        |     |
| TTG CAG GGT TGT TTT GTG GTG ACG TGC ACA CTG TGT GCA TTC ATC     | 478 |
| Leu Gln Gly Cys Phe Val Val Thr Cys Thr Leu Cys Ala Phe Ile     |     |
| 35 40 45                                                        |     |

## (2) INFORMATION FOR SEQ ID NO: 107:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 275 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Brain

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 133..273
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 87..227  
id W31692  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 45..121
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..77  
id W31692  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 123..273
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 76..226  
id H46855

est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 47..122
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..76  
id H46855  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 133..273
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 85..225  
id H49687  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 47..121
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..75  
id H49687  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 133..273
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 84..224  
id H50194  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 47..121
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..75  
id H50194  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 133..273
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 82..222  
id AA285085  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 50..122
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100

region 1..73  
id AA285085  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 153..191
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 4.9  
seq MLIMLGIFNVHS/AV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

```

CCCTGCGAGG GCATCCTGGG.CTTTCTCCCA CCGCTTTCCG AGCCCGCTTG CACCTCGGCG 60
ATCCCCGACT CCCTTCTTTA TGGCGTCGCT CCTGTGCTGT GGGCCGAAGC TGGCCGCCTG 120
CGGCATCGTG YRTCAGCGCC TGGGGAGTGA TC ATG TTG ATA ATG CTC GGA ATA 173
 Met Leu Ile Met Leu Gly Ile
 -10

TTT TTC AAT GTC CAT TCC GCT GTG TTG ATT GAG GAC GTT CCC TTC ACG 221
Phe Phe Asn Val His Ser Ala Val Leu Ile Glu Asp Val Pro Phe Thr
-5 1 5 10

GAG AAA GAT TTT GAG ANT GGC CCC CAG AAC ATA TAC AAC CTT TAC GAG 269
Glu Lys Asp Phe Glu Xaa Gly Pro Gln Asn Ile Tyr Asn Leu Tyr Glu
15 20 25

CAT GGG 275
His Gly

```

## (2) INFORMATION FOR SEQ ID NO: 108:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 350 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Normal prostate

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 82..223
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..142  
id W24852  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 231..320

(C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 92  
 region 150..239  
 id W24852  
 est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 256..321  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 90  
 region 1..66  
 id AA129007  
 est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 321..350  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 90  
 region 65..94  
 id AA129007  
 est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 9..344  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 4.5  
 seq AAVAVGMLXASYA/AV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

|                                                                  |     |
|------------------------------------------------------------------|-----|
| AGAGGGTT ATG GGA GGG CTC TGG CGT CCT GGA TGG AGG TGC GTT CCT TTC | 50  |
| Met Gly Gly Leu Trp Arg Pro Gly Trp Arg Cys Val Pro Phe          |     |
| -110 -105 -100                                                   |     |
| TGT GGC TGG CGC TGG ATC CAC CCT GGG TCT CCA ACC AGG GCT GCA GAG  | 98  |
| Cys Gly Trp Arg Trp Ile His Pro Gly Ser Pro Thr Arg Ala Ala Glu  |     |
| -95 -90 -85                                                      |     |
| AGG GTA GAG CCG TTT CTT AGG CCA GAG TGG AGT GGG ACA GGA GGT GCC  | 146 |
| Arg Val Glu Pro Phe Leu Arg Pro Glu Trp Ser Gly Thr Gly Gly Ala  |     |
| -80 -75 -70                                                      |     |
| GAG AGA GGA CTG AGG TGG CTT GGG ACA TGG AAG CGC TGC AGC CTT CGA  | 194 |
| Glu Arg Gly Leu Arg Trp Leu Gly Thr Trp Lys Arg Cys Ser Leu Arg  |     |
| -65 -60 -55                                                      |     |
| GCC CGG CAT CCA GCA TTG CAG CCG CCG CGG CGG CCT AAG AGC TCG AAC  | 242 |
| Ala Arg His Pro Ala Leu Gln Pro Pro Arg Arg Pro Lys Ser Ser Asn  |     |
| -50 -45 -40 -35                                                  |     |
| CCT TTC ACA CGC GCG SKV GAG GAG GAR CGG CGG CGG MAG AAC AAG ACG  | 290 |
| Pro Phe Thr Arg Ala Xaa Glu Glu Glu Arg Arg Arg Xaa Asn Lys Thr  |     |
| -30 -25 -20                                                      |     |
| ACC CTC ACT TAC GTG GCC GCT GTC GCC GTG GGC ATG CTN NGG GCG TCC  | 338 |
| Thr Leu Thr Tyr Val Ala Ala Val Ala Val Gly Met Leu Xaa Ala Ser  |     |
| -15 -10 -5                                                       |     |

TAC GCT GCC GTA  
Tyr Ala Ala Val  
1

(2) INFORMATION FOR SEQ ID NO: 109:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 419 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens  
(D) DEVELOPMENTAL STAGE: Fetal  
(F) TISSUE TYPE: kidney

(ix) FEATURE:

- [illegible]

(ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 132..248  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 4.2  
seq SDPLCVLFLNTSG/QQ

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

|                                                                 |                                                       |            |            |            |            |     |
|-----------------------------------------------------------------|-------------------------------------------------------|------------|------------|------------|------------|-----|
| AAATCCCTGC                                                      | GGTCCCAGCG                                            | TCGCTCCGGA | CGCTGCCAAC | CTGTTCTCCA | CCGTCGCTCG | 60  |
| ACTTCCACCT                                                      | CTAAGACTCC                                            | CACGAAACTC | AGGTTGAATA | ATTTCATCAA | TTACACAAC  | 120 |
| GAACTCAAGA                                                      | C ATG GCT GCC CAG TGT GTC ACA AAG GTG GCG CTG AAT GTT | 170        |            |            |            |     |
|                                                                 | Met Ala Ala Gln Cys Val Thr Lys Val Ala Leu Asn Val   |            |            |            |            |     |
|                                                                 | -35                                                   | -30        |            |            |            |     |
| TCC TGT GCC AAT CTT TTG GAT AAA GAT ATA GGG TCA AAG TCA GAC CCT | 218                                                   |            |            |            |            |     |
| Ser Cys Ala Asn Leu Leu Asp Lys Asp Ile Gly Ser Lys Ser Asp Pro |                                                       |            |            |            |            |     |
| -25                                                             | -20                                                   | -15        |            |            |            |     |
| TTA TGT GTG TTA TTT TTG AAT ACA AGT GGT CAA CAG TGG TAT GAG GTT | 266                                                   |            |            |            |            |     |
| Leu Cys Val Leu Phe Leu Asn Thr Ser Gly Gln Gln Trp Tyr Glu Val |                                                       |            |            |            |            |     |
| -10                                                             | -5                                                    | 1          | 5          |            |            |     |
| GAG CGC ACA GAA AGG ATT AAG AAT TGC TTG AAT CCC CAA TTT TCC AAG | 314                                                   |            |            |            |            |     |
| Glu Arg Thr Glu Arg Ile Lys Asn Cys Leu Asn Pro Gln Phe Ser Lys |                                                       |            |            |            |            |     |
| 10                                                              | 15                                                    | 20         |            |            |            |     |

|                                                                 |     |
|-----------------------------------------------------------------|-----|
| ACA TTT ATT ATT GAT TAC TAC TTT GAA GTG GTT CAG AAA TTG AAA TTT | 362 |
| Thr Phe Ile Ile Asp Tyr Tyr Phe Glu Val Val Gln Lys Leu Lys Phe |     |
| 25 30 35                                                        |     |
| GGG GTT TAT GAC ATC GRC AAC AAA ACT ATT GAG CTG AGT GAT GAT GAC | 410 |
| Gly Val Tyr Asp Ile Xaa Asn Lys Thr Ile Glu Leu Ser Asp Asp Asp |     |
| 40 45 50                                                        |     |
| TTC TTA GGG                                                     | 419 |
| Phe Leu Gly                                                     |     |
| 55                                                              |     |

## (2) INFORMATION FOR SEQ ID NO: 110:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 405 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (D) DEVELOPMENTAL STAGE: Fetal
- (F) TISSUE TYPE: kidney

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 63..402
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 35..374  
id W79829  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 77..377
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 1..301  
id H81957  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 373..404
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 90  
region 298..329  
id H81957  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 88..402
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99

region 2..316  
id H62624  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide  
(B) LOCATION: 85..294  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 3.7  
seq AVLDCAFYDPHTA/WS

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

```

AAGTGTCTCTG AGGGAAGCAA GGAGGCGGCG GCGGCCGCAG CGAGTGGCGA GTAGTGGAAA 60
CGTTGCTTCT GAGGGGAGCC CAAG ATG ACC GGT TCT AAC GAG TTC AAG CTG 111
 Met Thr Gly Ser Asn Glu Phe Lys Leu
 -70 -65

AAC CAG CCA CCC GAG GAT GGC ATC TCC TCC GTG AAG TTC AGC CCC AAC 159
Asn Gln Pro Pro Glu Asp Gly Ile Ser Ser Val Lys Phe Ser Pro Asn
-60 -55 -50

ACC TCC CAG TTC CTG CTT GTC TCC TCC TGG GAC ACG TCC GTG CGT CTC 207
Thr Ser Gln Phe Leu Leu Val Ser Ser Trp Asp Thr Ser Val Arg Leu
-45 -40 -35 -30

TAC GAT GTG CCG GCC AAC TCC ATG CGG CTC AAG TAC CAG CAC ACC GGC 255
Tyr Asp Val Pro Ala Asn Ser Met Arg Leu Lys Tyr Gln His Thr Gly
 -25 -20 -15

GCC GTC CTG GAC TGC GCC TTC TAC GAT CCA ACG CAT GCC TGG AGT GGA 303
Ala Val Leu Asp Cys Ala Phe Tyr Asp Pro Thr His Ala Trp Ser Gly
 -10 -5 1

GGA CTA GAT CAT CMV KTG AAA ATG CAT GAT TTG AAC ACT GAT CAA GAA 351
Gly Leu Asp His Xaa Xaa Lys Met His Asp Leu Asn Thr Asp Gln Glu
 5 10 15

AAT CTT GTT GGG ACC CAT GAT GCC CCT ATC AGA TGT GTT GAA TAC TGT 399
Asn Leu Val Gly Thr His Asp Ala Pro Ile Arg Cys Val Glu Tyr Cys
 20 25 30 35

CCA AGT
Pro Ser
405

```

## (2) INFORMATION FOR SEQ ID NO: 111:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 442 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Brain

(ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 48..365  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 99  
                           region 1..318  
                           id N31699  
                           est

(ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 365..420  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 100  
                           region 319..374  
                           id N31699  
                           est

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 299..373  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 3.6  
                           seq AHLWCWCGSHCCST/CV

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

```

AGTGTTCCT CAAATGGCGG TGTGAAGAGA GTTCGCCTGA GCCAGATCCC AGGTTTCACT 60
GAAGAACTT CTTAGAGATT CATTGCACCT CTGAGATTTA ATGTTTACAA CTTGGAGTTG 120
TCGACCTTCT TATAAGATAC ATTTTGAAG TCAAAATGAA AGTTTCTGT GAAGTTTtag 180
AAGAGTTATA CAAGAAGGTA CTTCTTGGAG CCACACTTGA AAATGACAGC CATGATTACG 240
TCTTTTATCT CAACCCAGCA GTTTCAGATC AAGATTGTTT TACAGCCACC TCCTTAGA 298
ATG GGC AAA CAC CTG TGG TAT CCA GGG CAG GCA TCA GCC CAT CTC TGT 346
Met Gly Lys His Leu Trp Tyr Pro Gly Gln Ala Ser Ala His Leu Cys
-25 -20 -15 -10

TGG TGT GGC TCC CAT TGC TGT AGC ACC TGT GTG TTT GAA GAC CAA CTC 394
Trp Cys Gly Ser His Cys Cys Ser Thr Cys Val Phe Glu Asp Gln Leu
 -5 1 5

TCA GAT GAG CGG TTC CAG AGA AGT AAT GCT CCT TCA GTT AAC AGT GAT 442
Ser Asp Glu Arg Phe Gln Arg Ser Asn Ala Pro Ser Val Asn Ser Asp
 10 15 20

```

(2) INFORMATION FOR SEQ ID NO: 112:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 391 base pairs  
 (B) TYPE: NUCLEIC ACID  
 (C) STRANDEDNESS: DOUBLE  
 (D) TOPOLOGY: LINEAR



(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (D) DEVELOPMENTAL STAGE: Fetal
- (F) TISSUE TYPE: kidney

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 81..386
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 3..308  
id T23663  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 81..386
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 3..308  
id T23653  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 90..386
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 1..297  
id T03538  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 126..342
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 1..217  
id H28147  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 356..386
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 233..263  
id H28147  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 144..363
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 1..225  
id R71352  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 173..211
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.5  
seq MLAVSLTVXLLGA/MM

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

```

AGTGAGGTGG TTTCTGCGGG TGAGGCTGGC GCCCGTACCA TGAGCGAGGC GGACGGGCTG 60
CGACAGCGCC GGCCCTGCG GCCCGCAAGT CGTCACAGAC GATGATGGCC AGGCCCCGGA 120
GGCTAAGGAC GGCAGCTCCT TTAGCGGCAG AGTTTCCGA GTGACCTTCT TG ATG CTG 178
 Met Leu

GCT GTT TCT CTC ACC GTT CBC CTG CTT GGA GCC ATG ATG CTG CTG GAA 226
Ala Val Ser Leu Thr Val Xaa Leu Leu Gly Ala Met Met Leu Leu Glu
-10 -5 1 5

TCT CCT ATA GAT CCA CAG CCT CTC AGC TTC AAA GAA CCC CCG CTC TTG 274
Ser Pro Ile Asp Pro Gln Pro Leu Ser Phe Lys Glu Pro Pro Leu Leu
 10 15 20

CTT GGT GTT CTG CAT CCA AAT ACG AAG CTG CGA CAG GCA GAA AGG CTG 322
Leu Gly Val Leu His Pro Asn Thr Lys Leu Arg Gln Ala Glu Arg Leu
 25 30 35

TTT GAA AAT CAA CTT GTT GGA CCG GAG TCC ATA GCA CAT ATT GGG GAT 370
Phe Glu Asn Gln Leu Val Gly Pro Glu Ser Ile Ala His Ile Gly Asp
 40 45 50

GTG ATG TTT ACT GGG AGC TGG 391
Val Met Phe Thr Gly Ser Trp
 55 60

```

## (2) INFORMATION FOR SEQ ID NO: 113:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 339 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Ovary

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 76..203
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..128  
id R57344

est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 204..235
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 128..159  
id R57344  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide
- (B) LOCATION: 82..309
- (C) IDENTIFICATION METHOD: Von Heijne matrix
- (D) OTHER INFORMATION: score 3.5  
seq MLELDLLVFHLWG/SQ

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

```

AAGTAGCGCC TGCWGGCGGY GGCAGTTTGC CCGCGGRWGT GTGAAGGGAG ACAGTGTGGA 60
GGCCACAGGG TACTCGCCAC G ATG AGC AGC ACC TTA GCT AAG ATC GCG GAG 111
 Met Ser Ser Thr Leu Ala Lys Ile Ala Glu
 -75 -70

ATA GAA GCA GAG ATG GCT CGG ACT CAA AAG AAC AAG GCC ACA GCA CAC 159
Ile Glu Ala Glu Met Ala Arg Thr Gln Lys Asn Lys Ala Thr Ala His
-65 -60 -55

CAC TTA GGG CTG CTT AAG GCT CGT CTT GCT AAG CTT CGT CGA GAA CTC 207
His Leu Gly Leu Leu Lys Ala Arg Leu Ala Lys Leu Arg Arg Glu Leu
-50 -45 -40 -35

ATT ACT CCA AAG GGT GGT GGT GGT GGA GGT CCA GGA GAA GGT TTT GAT 255
Ile Thr Pro Lys Gly Gly Gly Gly Gly Gly Pro Gly Glu Gly Phe Asp
 -30 -25 -20

TGG CCA AGA CAG GTG ATG CTC GAA TTG GAT TTG TTG GTT TTC CAT CTG 303
Trp Pro Arg Gln Val Met Leu Glu Leu Asp Leu Leu Val Phe His Leu
 -15 -10 -5

TGG GGA AGT CAA CAC TGC TTA GTA ACC TGG CAA GGG 339
Trp Gly Ser Gln His Cys Leu Val Thr Trp Gln Gly
 1 5 10

```

## (2) INFORMATION FOR SEQ ID NO: 114:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 217 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens

(D) DEVELOPMENTAL STAGE: Fetal  
(F) TISSUE TYPE: kidney

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 17..214  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 1..198  
id C18087  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 53..140  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 2..89  
id T73970  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 128..214  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 91  
region 76..162  
id T73970  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 93..140  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 44..91  
id T73946  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 60..142  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 92  
region 13..95  
id AA096472  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 144..173  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 96  
region 96..125  
id AA096472  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 169..214  
(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100  
region 1..46  
id AA280423  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 47..181  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 13.9  
seq LVLALLLVSAALS/SV

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

```

ATGGCGTAGA GCCTAGCAAC AGCGCAGGCT CCCAGCCGAG TCCGTT ATG GCC GCT 55
 Met Ala Ala
 -45

GCC GTC CCG AAG AGG ATG AGG GGG CCA GCA CAA GCG AAA CTG CTG CCC 103
Ala Val Pro Lys Arg Met Arg Gly Pro Ala Gln Ala Lys Leu Leu Pro
-40 -35 -30

GGG TCG GCC ATC CAA GCC CTT GTG GGG TTG GCG CGG CCG CTG GTC TTG 151
Gly Ser Ala Ile Gln Ala Leu Val Gly Leu Ala Arg Pro Leu Val Leu
-25 -20 -15

GCG CTC CTG CTT GTG TCC GCC GCT CTA TCC AGT GTT GTA TCA CGG ACT 199
Ala Leu Leu Leu Val Ser Ala Ala Leu Ser Ser Val Val Ser Arg Thr
-10 -5 1 5

GAT TCA CCG AGC CCA CTG 217
Asp Ser Pro Ser Pro Leu
10

```

## (2) INFORMATION FOR SEQ ID NO: 115:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 372 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens  
(D) DEVELOPMENTAL STAGE: Fetal  
(F) TISSUE TYPE: kidney

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 147..264  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 152..269  
id AA015703  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 316..366  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 94  
region 322..372  
id AA015703  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 257..302  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 261..306  
id AA015703  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 184..258  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 13.9  
seq LLSLLFLVQGAHG/RG

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

```
AACAAAGAGT TGGCAGATCA CGGATGGAGG GCAGCATCTC CCAACAGCCT GGGCGGCCGC 60
TGAGACCCAG AGAACCCAAG GACTCCCCTK GGGGGYWCA Y CCAGCAGCCT CTGCTTCCCA 120
GGAGAGAGGT GCTGAAGTCC ACGAAGAGGT GGTGACTTCC AAGAGTGA CTGTCGGAGG 180
AAA ATG ACT CCC CAG TCG CTG CTG CAG ACG ACA CTG TTC CTG CTG AGT 228
 Met Thr Pro Gln Ser Leu Leu Gln Thr Thr Leu Phe Leu Leu Ser
 -25 -20 -15

CTG CTC TTC CTG GTC CAA GGT GCC CAC GGC AGG GGC CAC AGG GAA GAC 276
Leu Leu Phe Leu Val Gln Gly Ala His Gly Arg Gly His Arg Glu Asp
-10 -5 1 5

TTT CGC TTC TGC AGC CAG CGG AAC CAG ACA CAC AGG AGC AGC CTC CAC 324
Phe Arg Phe Cys Ser Gln Arg Asn Gln Thr His Arg Ser Ser Leu His
 10 15 20

TAY AAA CCC ACA CCA GAM CTG CGC ATC TCC ATC GAG AAC TCC GAA GAG 372
Tyr Lys Pro Thr Pro Xaa Leu Arg Ile Ser Ile Glu Asn Ser Glu Glu
 25 30 35
```

## (2) INFORMATION FOR SEQ ID NO: 116:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 439 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Testis

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 36..390
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 43..397  
id W31335  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 2..34
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 10..42  
id W31335  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(151..440)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 64..353  
id N30852  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(82..157)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 348..423  
id N30852  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 51..314
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 1..264  
id HSPD03622  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 311..375
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 92  
region 262..326  
id HSPD03622  
est

## (ix) FEATURE:

- (A) NAME/KEY: other

(B) LOCATION: 389..434  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 93  
region 342..387  
id HSPD03622  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 2..316  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98  
region 9..323  
id AA055130  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 316..375  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 96  
region 324..383  
id AA055130  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 145..436  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 95  
region 123..414  
id H19862  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 50..110  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 95  
region 26..86  
id H19862  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 107..145  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 84..122  
id H19862  
est

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 59..322  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 11.6  
seq ILLCLLLALFASG/LI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:



|                                                                 |     |
|-----------------------------------------------------------------|-----|
| AACCCGGTTC AGCTCGCCTT TCTTGGCCAG AGGCGCCGGT TGGACTCACG GCGGGGGC | 58  |
| ATG ATG GTG GTG GGT ACG GGC ACC TCG CTG GCG CTC TCC TCC CTC CTG | 106 |
| Met Met Val Val Gly Thr Gly Thr Ser Leu Ala Leu Ser Ser Leu Leu |     |
| -85 -80 -75                                                     |     |
| TCC CTG CTG CTC TTT GCT GGG ATG CAG ATT TAC AGC CGT CAG CTG GCC | 154 |
| Ser Leu Leu Leu Phe Ala Gly Met Gln Ile Tyr Ser Arg Gln Leu Ala |     |
| -70 -65 -60                                                     |     |
| TCC ACC GAG TGG CTC ACC ATC CAG GGC GGC CTG CTT GGT TCG GGT CTC | 202 |
| Ser Thr Glu Trp Leu Thr Ile Gln Gly Gly Leu Leu Gly Ser Gly Leu |     |
| -55 -50 -45                                                     |     |
| TTC GTG TTC TCG CTC ACT GCC TTC AAT AAT CTG GAG AAT CTT GTC TTT | 250 |
| Phe Val Phe Ser Leu Thr Ala Phe Asn Asn Leu Glu Asn Leu Val Phe |     |
| -40 -35 -30 -25                                                 |     |
| GGC AAA GGA TTC CAA GCA AAG ATC TTC CCT GAG ATT CTC CTG TGC CTC | 298 |
| Gly Lys Gly Phe Gln Ala Lys Ile Phe Pro Glu Ile Leu Leu Cys Leu |     |
| -20 -15 -10                                                     |     |
| CTG TTG GCT CTC TTT GCA TCT GGC CTC ATC CAC CRA GTC TGT GTC ACC | 346 |
| Leu Leu Ala Leu Phe Ala Ser Gly Leu Ile His Xaa Val Cys Val Thr |     |
| -5 1 5                                                          |     |
| ACC TGC TTC ATC TTC TCC AGG GTT GGT CTG TAC TAC ATC AAC AAG ATC | 394 |
| Thr Cys Phe Ile Phe Ser Arg Val Gly Leu Tyr Tyr Ile Asn Lys Ile |     |
| 10 15 20                                                        |     |
| TCC TCC ACC CTG TAC CAG GCA GCA GCT CCA GTC CTC ACA CCA GCC     | 439 |
| Ser Ser Thr Leu Tyr Gln Ala Ala Ala Pro Val Leu Thr Pro Ala     |     |
| 25 30 35                                                        |     |

## (2) INFORMATION FOR SEQ ID NO: 117:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 457 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 11..74
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 93  
region 1..64  
id R86288  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 217..251  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 97  
 region 204..238  
 id T29670  
 est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 56..112  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 11.6  
 seq VFCLLAVAPGAHS/QE

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

|                                                                  |     |
|------------------------------------------------------------------|-----|
| ATCCAACAAC CACATCCCTT CTCTACAGAA GCCTCTGAGA AGAAAGTTCT TCACC ATG | 58  |
| Met                                                              |     |
| GAC TGG ACC TGG AGG GTC TTC TGC TTG CTG GCT GTA GCT CCA GGT GCT  | 106 |
| Asp Trp Thr Trp Arg Val Phe Cys Leu Leu Ala Val Ala Pro Gly Ala  |     |
| -15 -10 -5                                                       |     |
| CAC TCC CAG GAA CAA CTG GTG CAG TCT GGG GCT GAG GTG TTG AAG CCT  | 154 |
| His Ser Gln Glu Gln Leu Val Gln Ser Gly Ala Glu Val Leu Lys Pro  |     |
| 1 5 10                                                           |     |
| GGG GCC TCA GTG AAC ATT TCC TGC AGG GCA TCT GGG TTC ACC TTC ACC  | 202 |
| Gly Ala Ser Val Asn Ile Ser Cys Arg Ala Ser Gly Phe Thr Phe Thr  |     |
| 15 20 25 30                                                      |     |
| AAT TAT TAT GTG CAC TGG GTG CGA CAG GCC CCT GGA CAC GGG CTT GAG  | 250 |
| Asn Tyr Tyr Val His Trp Val Arg Gln Ala Pro Gly His Gly Leu Glu  |     |
| 35 40 45                                                         |     |
| TGG ATG GGA GTG ATC AAC CCC GTT AGT GGT TAC ACA AGT TAC GCA CAG  | 298 |
| Trp Met Gly Val Ile Asn Pro Val Ser Gly Tyr Thr Ser Tyr Ala Gln  |     |
| 50 55 60                                                         |     |
| AAA CTG CAG GGC AGA CTG ACC ATG ACC ACG GAC ACG GCC GCG AAT ATA  | 346 |
| Lys Leu Gln Gly Arg Leu Thr Met Thr Thr Asp Thr Ala Ala Asn Ile  |     |
| 65 70 75                                                         |     |
| GTC TAC ATG GAC CTC AGT AGG CTG AAA TCT GAC GAC ACG GCC GTG TAT  | 394 |
| Val Tyr Met Asp Leu Ser Arg Leu Lys Ser Asp Asp Thr Ala Val Tyr  |     |
| 80 85 90                                                         |     |
| TTC TGT GCG AAA GTG CGG TGT CTT AAG GGG ATA TGC TAT ACA GAG GAT  | 442 |
| Phe Cys Ala Lys Val Arg Cys Leu Lys Gly Ile Cys Tyr Thr Glu Asp  |     |
| 95 100 105 110                                                   |     |
| GCT CTG GAT CTT TGG                                              | 457 |
| Ala Leu Asp Leu Trp                                              |     |
| 115                                                              |     |

## (2) INFORMATION FOR SEQ ID NO: 118:

- (i) SEQUENCE CHARACTERISTICS:  
    (A) LENGTH: 439 base pairs  
    (B) TYPE: NUCLEIC ACID  
    (C) STRANDEDNESS: DOUBLE  
    (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:  
    (A) ORGANISM: Homo Sapiens  
    (F) TISSUE TYPE: Umbilical cord
- (ix) FEATURE:  
    (A) NAME/KEY: other  
    (B) LOCATION: 75..429  
    (C) IDENTIFICATION METHOD: blastn  
    (D) OTHER INFORMATION: identity 99  
                            region 43..397  
                            id W31335  
                            est
- (ix) FEATURE:  
    (A) NAME/KEY: other  
    (B) LOCATION: 32..73  
    (C) IDENTIFICATION METHOD: blastn  
    (D) OTHER INFORMATION: identity 97  
                            region 1..42  
                            id W31335  
                            est
- (ix) FEATURE:  
    (A) NAME/KEY: other  
    (B) LOCATION: 33..355  
    (C) IDENTIFICATION METHOD: blastn  
    (D) OTHER INFORMATION: identity 98  
                            region 1..323  
                            id AA055130  
                            est
- (ix) FEATURE:  
    (A) NAME/KEY: other  
    (B) LOCATION: 355..414  
    (C) IDENTIFICATION METHOD: blastn  
    (D) OTHER INFORMATION: identity 100  
                            region 324..383  
                            id AA055130  
                            est
- (ix) FEATURE:  
    (A) NAME/KEY: other  
    (B) LOCATION: 56..384  
    (C) IDENTIFICATION METHOD: blastn  
    (D) OTHER INFORMATION: identity 96  
                            region 28..356  
                            id AA252648  
                            est
- (ix) FEATURE:  
    (A) NAME/KEY: other  
    (B) LOCATION: 385..428  
    (C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100  
region 356..399  
id AA252648  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 113..439  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 38..364  
id AA228934  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 184..440  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 123..379  
id H19862  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 89..149  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 90  
region 26..86  
id H19862  
est

(ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 146..184  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 84..122  
id H19862  
est

(ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 23..361  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 11.6  
seq ILLCLLLALFASG/LI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

|                                                                  |     |
|------------------------------------------------------------------|-----|
| AAGTCCGCGG TAAGGCTGAC GC ATG CGC ATA GCT AAC CGC ACC CGG TTC AGC | 52  |
| Met Arg Ile Ala Asn Arg Thr Arg Phe Ser                          |     |
| -110 -105                                                        |     |
| TCG CCT TTC TTG GCC AGA GGC GCC GGT TGG ACT CAC GGG CGG GGC ATG  | 100 |
| Ser Pro Phe Leu Ala Arg Gly Ala Gly Trp Thr His Gly Arg Gly Met  |     |
| -100 -95 -90                                                     |     |
| ATG GTG GTG GGT ACG GGC ACC TCG CTG GCG CTC TSS TCC CTC CTG TCC  | 148 |
| Met Val Val Gly Thr Gly Thr Ser Leu Ala Leu Xaa Ser Leu Leu Ser  |     |

| -85                                                             | -80 | -75 |     |
|-----------------------------------------------------------------|-----|-----|-----|
| CTG CTG CTC TTT GCT GGG ATG CAG ATG TAC AGC CGT CAG CTG GCC TCC |     |     | 196 |
| Leu Leu Leu Phe Ala Gly Met Gln Met Tyr Ser Arg Gln Leu Ala Ser |     |     |     |
| -70                                                             | -65 | -60 |     |
| ACC GAG TGG CTC ACC ATC CAG GGC GGC CTG CTT GGT TCG GGT CTC TTC |     |     | 244 |
| Thr Glu Trp Leu Thr Ile Gln Gly Gly Leu Leu Gly Ser Gly Leu Phe |     |     |     |
| -55                                                             | -50 | -45 | -40 |
| GTG TTC TCG CTC ACT GCC TTC AAT AAT CTG GAG AAT CTT GTC TTT GGC |     |     | 292 |
| Val Phe Ser Leu Thr Ala Phe Asn Asn Leu Glu Asn Leu Val Phe Gly |     |     |     |
| -35                                                             | -30 | -25 |     |
| AAA GGA TTC CAA GCA AAG ATC TTC CCT GAG ATT CTC CTG TGC CTC CTG |     |     | 340 |
| Lys Gly Phe Gln Ala Lys Ile Phe Pro Glu Ile Leu Leu Cys Leu Leu |     |     |     |
| -20                                                             | -15 | -10 |     |
| TTG GCT CTC TTT GCA TCT GGC CTC ATC CAC CGA GTC TGT GTC ACC ACC |     |     | 388 |
| Leu Ala Leu Phe Ala Ser Gly Leu Ile His Arg Val Cys Val Thr Thr |     |     |     |
| -5                                                              | 1   | 5   |     |
| TGC TTC ATC TTC TCC ATG GTT GGT CTG TAC TAC ATC AAC AAG ATC TCC |     |     | 436 |
| Cys Phe Ile Phe Ser Met Val Gly Leu Tyr Tyr Ile Asn Lys Ile Ser |     |     |     |
| 10                                                              | 15  | 20  | 25  |
| TCC                                                             |     |     | 439 |
| Ser                                                             |     |     |     |

## (2) INFORMATION FOR SEQ ID NO: 119:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 309 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Heart

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 16..250
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 1..235  
id AA280774  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 246..282
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 230..266  
id AA280774

est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 17..259  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 1..243  
id HUM404F03B  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 20..282  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 1..263  
id W05476  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 21..282  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 1..262  
id R33542  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 12..282  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 100  
region 8..278  
id T85491  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 151..222  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 11.4  
seq LMSLLLVPVVEA/VE

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

```
ADTCCTGTAA TGGCTGCTTC CTAGAAGGTC GTGTCACGTG GAACCTCTTA ATCTCAGCAT 60
CCGGAGCTCC AGGAAGGGAA AATTTC AAGT CAGATAGAAT TCTATATATA CCATTCTTT 120
GGAACCTTCA GCCCTCAAGA TTCCAACATC ATG ACC TCA GTT TCA ACA CAG TTG 174
 Met Thr Ser Val Ser Thr Gln Leu
 -20
TCC TTA GTC CTC ATG TCA CTG CTT TTG GTG CTG CCT GTT GTG GAA GCA 222
Ser Leu Val Leu Met Ser Leu Leu Val Leu Pro Val Val Glu Ala
-15 -10 -5
GTA GAA GCC GGT GAT GCA ATC GCC CTT TTG TTA GGT GTG GTT CTC AGC 270
```

Val Glu Ala Gly Asp Ala Ile Ala Leu Leu Leu Gly Val Val Leu Ser  
1 5 10 15

ATT ACA GGC ATT GTG CCT GCT TGG GGG TAT ATG CAY GGG 309  
Ile Thr Gly Ile Val Pro Ala Trp Gly Tyr Met His Gly  
20 25

(2) INFORMATION FOR SEQ ID NO: 120:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 361 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Brain

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 95..363
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 60..328  
id H19572  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 140..290
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 106..256  
id H46195  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 95..148
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 92  
region 62..115  
id H46195  
est

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(207..316)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 183..292  
id H46196  
est

(ix) FEATURE:

- (A) NAME/KEY: other

(B) LOCATION: complement(314..363)  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98  
region 137..186  
id H46196  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: complement(172..212)  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 95  
region 288..328  
id H46196  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: complement(237..287)  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 92  
region 239..289  
id H19490  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: complement(284..317)  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 97  
region 208..241  
id H19490  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: complement(331..363)  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 93  
region 160..192  
id H19490  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 263..322  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 11.2  
seq ILVVLMLPLAQA/LD

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

AAGACACGCC TACGATTAGA CTCAGGCAGG CACCTACCGG CGAGCGGCCG CRVGTGACTC 60  
CCAGGCGCGG CGGTACCTCA CGGTGGTGAA GGTACAGGG TTGCAGCACT CCCAGTAGAC 120  
CAGGAGCTCC GGSAGGCAGG GCCGGCCCCA CGTCCTCTGC GCACCACCCT GAGTTGGATC 180  
CTCTGTGCGC CACCCCTGAG TTGGATCCAG GGCTAGCTGC TGTTGACCTC CCCACTCCCA 240



```
CGCTGCCCTC CTGCCTGCAG CC ATG ACG CCC CTG CTC ACC CTG ATC CTG GTG 292
 Met Thr Pro Leu Leu Thr Leu Ile Leu Val
 -20 -15

GTC CTC ATG GGC TTA CCT CTG GCC CAG GCC TTG GAC TGC CAC GTG TGT 340
Val Leu Met Gly Leu Pro Leu Ala Gln Ala Leu Asp Cys His Val Cys
-10 -5 1 5

NCC TAC AAC GGA GAC AAC TGC 361
Xaa Tyr Asn Gly Asp Asn Cys
 10
```

## (2) INFORMATION FOR SEQ ID NO: 121:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 510 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Lymph ganglia

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 20..372
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 1..353  
id W05519  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 368..423
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 348..403  
id W05519  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 17..260
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 21..264  
id T97490  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 231..341
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 287..347

id T97490  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 16..315  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 99  
region 1..300  
id HUML12811  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 16..275  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 98  
region 1..260  
id HUML13801  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 139..186  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 11  
seq LLALSLLVLWTSP/AP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

```
AATCCCAGC CTCACATCAC TCACACCTTG CATTTCACCC CTGCATCCCA GTCGCCCTGC 60
AGCCTCACAC AGATCCTGCA CACACCCAGA CAGCTGGCGC TCACACATTC ACCGTTGGCC 120
TGCCTCTGTT CACCCTCC ATG GCC CTG CTA CTG GCC CTC AGC CTG CTA GTT 171
 Met Ala Leu Leu Leu Ala Leu Ser Leu Leu Val
 -15 -10

CTC TGG ACT TCC CCA GCC CCA ACT CTG AGT GGC ACC AAT GAT GCT GAA 219
Leu Trp Thr Ser Pro Ala Pro Thr Leu Ser Gly Thr Asn Asp Ala Glu
-5 1 5 10

GAC TGC TGC CTG TCT GTG ACC CAG AAA CCC ATC CCT GGG TAC ATC GTG 267
Asp Cys Cys Leu Ser Val Thr Gln Lys Pro Ile Pro Gly Tyr Ile Val
 15 20 25

AGG AAC TTC CAC TAC CTT CTC ATC AAG GAT GGC TGC AGG GTG CCT GCT 315
Arg Asn Phe His Tyr Leu Leu Ile Lys Asp Gly Cys Arg Val Pro Ala
 30 35 40

GTA GTG TTC ACC ACA CTG AGG GGC CGC CAG CTC TGT GCA CCC CCA GAC 363
Val Val Phe Thr Thr Leu Arg Gly Arg Gln Leu Cys Ala Pro Pro Asp
 45 50 55

CAG CCC TGG GTA GAA CGC ATC ATC CAG AGA CTG CAG AGG ACC TCA GCC 411
Gln Pro Trp Val Glu Arg Ile Ile Gln Arg Leu Gln Arg Thr Ser Ala
 60 65 70 75

AAG ATG AAR MGC CGM AGC AGT KAA CCT ATG AMC GTG MAG AGG GAR CCG 459
Lys Met Lys Xaa Arg Ser Ser Xaa Pro Met Xaa Val Xaa Arg Glu Pro
```

|                                                                 | 80 | 85  | 90  |     |
|-----------------------------------------------------------------|----|-----|-----|-----|
| GAG TCC GAG TCA AGC ATT GTG AAT KAT TAC CTA MCT GGG GAA CGA RGA |    |     |     | 507 |
| Glu Ser Glu Ser Ser Ile Val Asn Xaa Tyr Leu Xaa Gly Glu Arg Xaa |    |     |     |     |
|                                                                 | 95 | 100 | 105 |     |
| AGG                                                             |    |     |     | 510 |
| Arg                                                             |    |     |     |     |

## (2) INFORMATION FOR SEQ ID NO: 122:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 382 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 152..287
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 91..226  
id W60940  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 108..160
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 48..100  
id W60940  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 60..106
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..47  
id W60940  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 152..316
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 90..254  
id H39980  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 62..160
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..99  
id H39980  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 308..384
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 247..323  
id H39980  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(148..292)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 99  
region 282..426  
id N41026  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(283..384)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 93  
region 191..292  
id N41026  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 66..160
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 8..102  
id R49793  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 199..271
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 141..213  
id R49793  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 152..199
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 91  
region 93..140  
id R49793

est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 18..160  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 96  
 region 1..143  
 id W74783  
 est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 190..253  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 98  
 region 173..236  
 id W74783  
 est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 74..136  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 10.5  
 seq RLLLLPLLLAVSG/LR

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

```

AATTTCACTT GCCTGGACGC TGCGCCACAT CCCACCGGCC CTTACTGTGT GGTGTCCAGC 60
AGCATCCGGC TTC ATG GGG GGA CTT GAA CCC TGC AGC AGG CTC CTG CTC 109
 Met Gly Gly Leu Glu Pro Cys Ser Arg Leu Leu Leu
 -20 -15 -10

CTG CCT CTC CTG CTG GCT GTA AGT GGT CTC CGT CCT GTC CAG GCC CAG 157
Leu Pro Leu Leu Ala Val Ser Gly Leu Arg Pro Val Gln Ala Gln
 -5 1 5

GCC CAG AGC GAT TGC AGT TGC TCT ACG GTG AGC CCG GGC GTG CTG GCA 205
Ala Gln Ser Asp Cys Ser Cys Ser Thr Val Ser Pro Gly Val Leu Ala
 10 15 20

GGG ATC GTG ATG GGA GAC CTG GTG CTG ACA GTG CTC ATT GCC CTG GCC 253
Gly Ile Val Met Gly Asp Leu Val Leu Thr Val Leu Ile Ala Leu Ala
 25 30 35

GTG TAC TTC CTG GGC CGG CTG GTC CCT CGG GGG CGA GGG GCT GCG GAG 301
Val Tyr Phe Leu Gly Arg Leu Val Pro Arg Gly Arg Gly Ala Ala Glu
 40 45 50 55

GCA SNG ACC CGG AAA CAG CGT ATC ACT GAG ACC GGG TCG CCT TAT CAG 349
Ala Xaa Thr Arg Lys Gln Arg Ile Thr Glu Thr Gly Ser Pro Tyr Gln
 60 65 70

GAG CTC CAG GGT CAG AGG TCG GAT GTC TAC AGC 382
Glu Leu Gln Gly Gln Arg Ser Asp Val Tyr Ser
 75 80

```

## (2) INFORMATION FOR SEQ ID NO: 123:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 423 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (D) DEVELOPMENTAL STAGE: Fetal
- (F) TISSUE TYPE: kidney

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 54..196
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 13..155  
id N41450  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 193..332
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 92  
region 153..292  
id N41450  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 327..425
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 288..386  
id N41450  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 204..332
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 202..330  
id W76359  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 54..124
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 54..124  
id W76359  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 2..53
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 3..54  
id W76359  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 327..370
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 326..369  
id W76359  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 164..196
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 162..194  
id W76359  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 133..163
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 132..162  
id W76359  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 54..128
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 46..120  
id W04321  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 9..54
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 2..47  
id W04321  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 164..201
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 94  
region 153..190  
id W04321

est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 125..163
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 92  
region 115..153  
id W04321  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 2..124
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 12..134  
id AA025985  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 200..286
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 95  
region 208..294  
id AA025985  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 366..425
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 381..440  
id AA025985  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 135..166
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 145..176  
id AA025985  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 208..306
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 146..244  
id H09017  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 62..126
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98



region 1..65  
id H09017  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
(B) LOCATION: 327..368  
(C) IDENTIFICATION METHOD: blastn  
(D) OTHER INFORMATION: identity 92  
region 267..308  
id H09017  
est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
(B) LOCATION: 178..249  
(C) IDENTIFICATION METHOD: Von Heijne matrix  
(D) OTHER INFORMATION: score 10  
seq LCRALCLFPRVFA/AE

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

```

AAAGGACTCC AAAGCGAGGC CGGGGACTGA AGGTGTGGGT GTCGAGCCCT CTGGCAGAGG 60
GTTAACCTGG GTCAAATGCA CGGATTCTCA CCTCGTACAG TTACGCTCTC CCGCGGCACG 120
TCCGCGAGGA CITGAAGTCC TGAGCGCTCA AGTTTGTCCG TAGGTCGAGA GAAGGCC 177
ATG GAG GTG CCG CCA CCG GCA CCG CGG AGC TTT CTC TGT AGA GCA TTG 225
Met Glu Val Pro Pro Ala Pro Arg Ser Phe Leu Cys Arg Ala Leu
 -20 -15 -10
TGC CTA TTT CCC CGA GTC TTT GCT GCC GAA GCT GTG ACT GCC GAT TCG 273
Cys Leu Phe Pro Arg Val Phe Ala Ala Glu Ala Val Thr Ala Asp Ser
 -5 1 5
GAA GTC CTT GAG GAG CGT CAG AAG CGG CTT CCC TAC STC CCA GAG CCC 321
Glu Val Leu Glu Glu Arg Gln Lys Arg Leu Pro Tyr Xaa Pro Glu Pro
 10 15 20
TAT TAC CGG AAT CTG GAT GGG ACC GCC TCC GGG AGC TGT TTK GCA AAG 369
Tyr Tyr Arg Asn Leu Asp Gly Thr Ala Ser Gly Ser Cys Xaa Ala Lys
 25 30 35 40
ATG AAC AGC AGA GAA TTT CAA AGG ACC TTG CTA ATA TCT GTA AGA CGG 417
Met Asn Ser Arg Glu Phe Gln Arg Thr Leu Leu Ile Ser Val Arg Arg
 45 50 55
CAG CTA
Gln Leu 423

```

## (2) INFORMATION FOR SEQ ID NO: 124:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 356 base pairs  
(B) TYPE: NUCLEIC ACID  
(C) STRANDEDNESS: DOUBLE  
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo Sapiens

(F) TISSUE TYPE: Cancerous prostate

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 8..208

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100  
region 1..201  
id N56128  
est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 242..311

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 94  
region 233..302  
id N56128  
est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 207..244

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 94  
region 199..236  
id N56128  
est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 19..113

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 100  
region 1..95  
id N87312  
est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 223..286

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 90  
region 208..271  
id N87312  
est

(ix) FEATURE:

(A) NAME/KEY: other

(B) LOCATION: 181..222

(C) IDENTIFICATION METHOD: blastn

(D) OTHER INFORMATION: identity 92  
region 165..206  
id N87312  
est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 46..270  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 97  
 region 1..225  
 id R57616  
 est

## (ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: 51..241  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 99  
 region 1..191  
 id AA093451  
 est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 75..131  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 9.5  
 seq LMCLSLCTAFALS/KP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

```

AGAGCTGAGC CGGTGGGTGA GCGGCGGCCA CGGCATCCTG TGCTGTGGGG GCTACGAGGA 60
AAGATCTAAT TATC ATG GAC CTG CGA CAG TTT CTT ATG TGC CTG TCC CTG 110
 Met Asp Leu Arg Gln Phe Leu Met Cys Leu Ser Leu
 -15 -10
TGC ACA GCC TTT GCC TTG AGC AAA CCC ACA GAA AAG AAG GAC CGT GTA 158
Cys Thr Ala Phe Ala Leu Ser Lys Pro Thr Glu Lys Lys Asp Arg Val
 -5 1 5
CAT CAT GAG CCT CAG CTC AGT GAC AAG GTT CAC AAT GAT GCT CAG AGT 206
His His Glu Pro Gln Leu Ser Asp Lys Val His Asn Asp Ala Gln Ser
 10 15 20 25
TTT GWT TAT GAC CAT GAT GCC TTC TTG GGT GCT GAA GAA GCA AAG ASM 254
Phe Xaa Tyr Asp His Asp Ala Phe Leu Gly Ala Glu Glu Ala Lys Xaa
 30 35 40
TTT GAT CAG CTG ACA CCA GAA GAG AGC AAG GAA AGG CTT GGA AAG ATT 302
Phe Asp Gln Leu Thr Pro Glu Glu Ser Lys Glu Arg Leu Gly Lys Ile
 45 50 55
GTA AGT AAR ATM GAT GGC GAC AAG GAC GGG TTT GTC ACT GTG GAT GAG 350
Val Ser Lys Ile Asp Gly Asp Lys Asp Gly Phe Val Thr Val Asp Glu
 60 65 70
CTC AAA
Leu Lys
 75

```

## (2) INFORMATION FOR SEQ ID NO: 125:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 320 base pairs
  - (B) TYPE: NUCLEIC ACID
  - (C) STRANDEDNESS: DOUBLE
  - (D) TOPOLOGY: LINEAR
- (ii) MOLECULE TYPE: CDNA
- (vi) ORIGINAL SOURCE:
  - (A) ORGANISM: Homo Sapiens
  - (F) TISSUE TYPE: Substantia nigra
- (ix) FEATURE:
  - (A) NAME/KEY: other
  - (B) LOCATION: 50..320
  - (C) IDENTIFICATION METHOD: blastn
  - (D) OTHER INFORMATION: identity 96  
region 17..287  
id R35366  
est
- (ix) FEATURE:
  - (A) NAME/KEY: other
  - (B) LOCATION: 42..320
  - (C) IDENTIFICATION METHOD: blastn
  - (D) OTHER INFORMATION: identity 97  
region 10..288  
id R35909  
est
- (ix) FEATURE:
  - (A) NAME/KEY: other
  - (B) LOCATION: 42..318
  - (C) IDENTIFICATION METHOD: blastn
  - (D) OTHER INFORMATION: identity 97  
region 10..286  
id R20566  
est
- (ix) FEATURE:
  - (A) NAME/KEY: other
  - (B) LOCATION: 42..320
  - (C) IDENTIFICATION METHOD: blastn
  - (D) OTHER INFORMATION: identity 96  
region 10..288  
id H09254  
est
- (ix) FEATURE:
  - (A) NAME/KEY: other
  - (B) LOCATION: 42..320
  - (C) IDENTIFICATION METHOD: blastn
  - (D) OTHER INFORMATION: identity 96  
region 10..283  
id R25274  
est
- (ix) FEATURE:
  - (A) NAME/KEY: sig\_peptide
  - (B) LOCATION: 24..113

(C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 9.5  
 seq LLFLSQFCILSGG/ES

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

```

AAAAGTGGCG AGGCGCTGGC AAG ATG GCG GGA GGG GTG CGC CCG CTG CGG GGC 53
 Met Ala Gly Gly Val Arg Pro Leu Arg Gly
 -30 -25

CTC CGC GCC TTG TGT CGC GTG CTG CTC TTC CTC TCG CAG TTC TGC ATT 101
Leu Arg Ala Leu Cys Arg Val Leu Leu Phe Leu Ser Gln Phe Cys Ile
-20 -15 -10 -5

CTG TCG GGC GGT GAA AGT ACT GAA ATC CCA CCT TAT GTG ATG AAG TGT 149
Leu Ser Gly Gly Glu Ser Thr Glu Ile Pro Pro Tyr Val Met Lys Cys
 1 5 10

CCG AGC AAT GGT TTG TGT AGC AGG CTT CCT GCA GAC TGT ATA GAC AGC 197
Pro Ser Asn Gly Leu Cys Ser Arg Leu Pro Ala Asp Cys Ile Asp Ser
 15 20 25

ACA ACA AAT TTC TCC TGT ACC TAT GGG AAG CCT GTM ACT TTT GAC TGT 245
Thr Thr Asn Phe Ser Cys Thr Tyr Gly Lys Pro Val Thr Phe Asp Cys
 30 35 40

RCA GTG AAA CCA TCT GTT ACC TGT GTT GAT CAA GAC TTC AAA TCC CAA 293
Xaa Val Lys Pro Ser Val Thr Cys Val Asp Gln Asp Phe Lys Ser Gln
 45 50 55 60

AAG RAC TTC ATC ATT AAC ATG ACT TGC 320
Lys Xaa Phe Ile Ile Asn Met Thr Cys
 65

```

(2) INFORMATION FOR SEQ ID NO: 126:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 389 base pairs
- (B) TYPE: NUCLEIC ACID
- (C) STRANDEDNESS: DOUBLE
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Umbilical cord

(ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: complement(2..198)
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 93  
region 2..198  
id N27605  
est

(ix) FEATURE:

(A) NAME/KEY: other  
 (B) LOCATION: complement(2..69)  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 95  
 region 1..68  
 id N78549  
 est

## (ix) FEATURE:

(A) NAME/KEY: sig\_peptide  
 (B) LOCATION: 36..98  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 9.3  
 seq VLPVILLLLGAHP/SP

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

```

AAAATGCTTT CGGTAGGCAC TCCAMGGCTG TRAAG ATG GCG GCG GCT GCG TGG 53
 Met Ala Ala Ala Ala Trp
 -20

CTT CAG GTG TTG CCT GTC ATT CTT CTG CTT CTG GGA GCT CAC CCG TCA 101
Leu Gln Val Leu Pro Val Ile Leu Leu Leu Leu Gly Ala His Pro Ser
-15 -10 -5 1

CCA CTG TCG TTT TTC AGT GCG GGA CCG GCA ACC GTA GCT GCT GCC GAC 149
Pro Leu Ser Phe Phe Ser Ala Gly Pro Ala Thr Val Ala Ala Ala Asp
 5 10 15

CGG TCC AAA TGG CAC ATT CCG ATA CCG TCG GGG AAA AAT TAT TTT AGT 197
Arg Ser Lys Trp His Ile Pro Ile Pro Ser Gly Lys Asn Tyr Phe Ser
 20 25 30

TTT GGA AHK ATC CTC TTC AGA AAT ACC ACT ATC TTC CTG AAG TTT GAT 245
Phe Gly Xaa Ile Leu Phe Arg Asn Thr Thr Ile Phe Leu Lys Phe Asp
 35 40 45

GGA GAA CCT TGT GAC CTG TCT TTG AAT ATA AYM TGG TAT CTG AAA AGC 293
Gly Glu Pro Cys Asp Leu Ser Leu Asn Ile Xaa Trp Tyr Leu Lys Ser
 50 55 60 65

GCT GAT TGT TAC AAT GAA ATC TAT AAC TTC AAG GCA GAA GAA GTA GAG 341
Ala Asp Cys Tyr Asn Glu Ile Tyr Asn Phe Lys Ala Glu Glu Val Glu
 70 75 80

TTG TAT TTG GAA AAA CTT AAG GAA AAA AGA GGC TTG TCT GGG AAA TGG 389
Leu Tyr Leu Glu Lys Leu Lys Glu Lys Arg Gly Leu Ser Gly Lys Trp
 85 90 95

```

## (i) INFORMATION FOR SEQ ID NO: 127:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 304 base pairs  
 (B) TYPE: NUCLEIC ACID  
 (C) STRANDEDNESS: DOUBLE  
 (D) TOPOLOGY: LINEAR

## (ii) MOLECULE TYPE: CDNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Homo Sapiens
- (F) TISSUE TYPE: Substantia nigra

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 31..297
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 96  
region 1..267  
id HSC1WH101  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 134..297
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 97  
region 41..204  
id R12437  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 95..136
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 1..42  
id R12437  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 95..297
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 98  
region 1..203  
id R13448  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 244..297
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 92  
region 82..135  
id T69236  
est

## (ix) FEATURE:

- (A) NAME/KEY: other
- (B) LOCATION: 197..244
- (C) IDENTIFICATION METHOD: blastn
- (D) OTHER INFORMATION: identity 100  
region 36..83  
id T69236  
est

## (ix) FEATURE:

- (A) NAME/KEY: sig\_peptide

(B) LOCATION: 212..268  
 (C) IDENTIFICATION METHOD: Von Heijne matrix  
 (D) OTHER INFORMATION: score 9.3  
 seq LLWLALACSPVHT/XL

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

```

ATCCGGCGCG CTGGAGCGTT TTCCGGCCGT GCGTTTGTGG CCGTCCGGCC TCCCTGACAT 60
GCAGATTTCC ANSSAGAAGA CAGAGAAGGA GCNAGTGGTC ATGGAATGGG CTGGGGTCAA 120
AGACTGGGTG CCTGGGAGCT GAGGCAGCCA CCGTTTCAGC CTGGCCAGCC CTCTGGACCC 180
CGAGGTTGGA CCCTACTGTG ACACACCTAC C ATG CGG ACA CTC TTC AAC CTC 232
 Met Arg Thr Leu Phe Asn Leu
 -15

CTC TGG CTT GCC CTG GCC TGC AGC CCT GTT CAC ACT ASC CTG TCA AAG 280
Leu Trp Leu Ala Leu Ala Cys Ser Pro Val His Thr Xaa Leu Ser Lys
-10 -5 1

TCA GAT GCC VSA AAA CCG CCT AGG 304
Ser Asp Ala Xaa Lys Pro Pro Arg
5 10

```

(2) INFORMATION FOR SEQ ID NO: 128:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 216 base pairs  
 (B) TYPE: NUCLEIC ACID  
 (C) STRANDEDNESS: DOUBLE  
 (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: CDNA

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: Homo Sapiens  
 (F) TISSUE TYPE: Muscle

(ix) FEATURE:  
 (A) NAME/KEY: other  
 (B) LOCATION: 43..162  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 99  
 region 29..148  
 id T98462  
 est

(ix) FEATURE:  
 (A) NAME/KEY: other  
 (B) LOCATION: 179..216  
 (C) IDENTIFICATION METHOD: blastn  
 (D) OTHER INFORMATION: identity 92  
 region 165..202  
 id T98462  
 est

(ix) FEATURE: